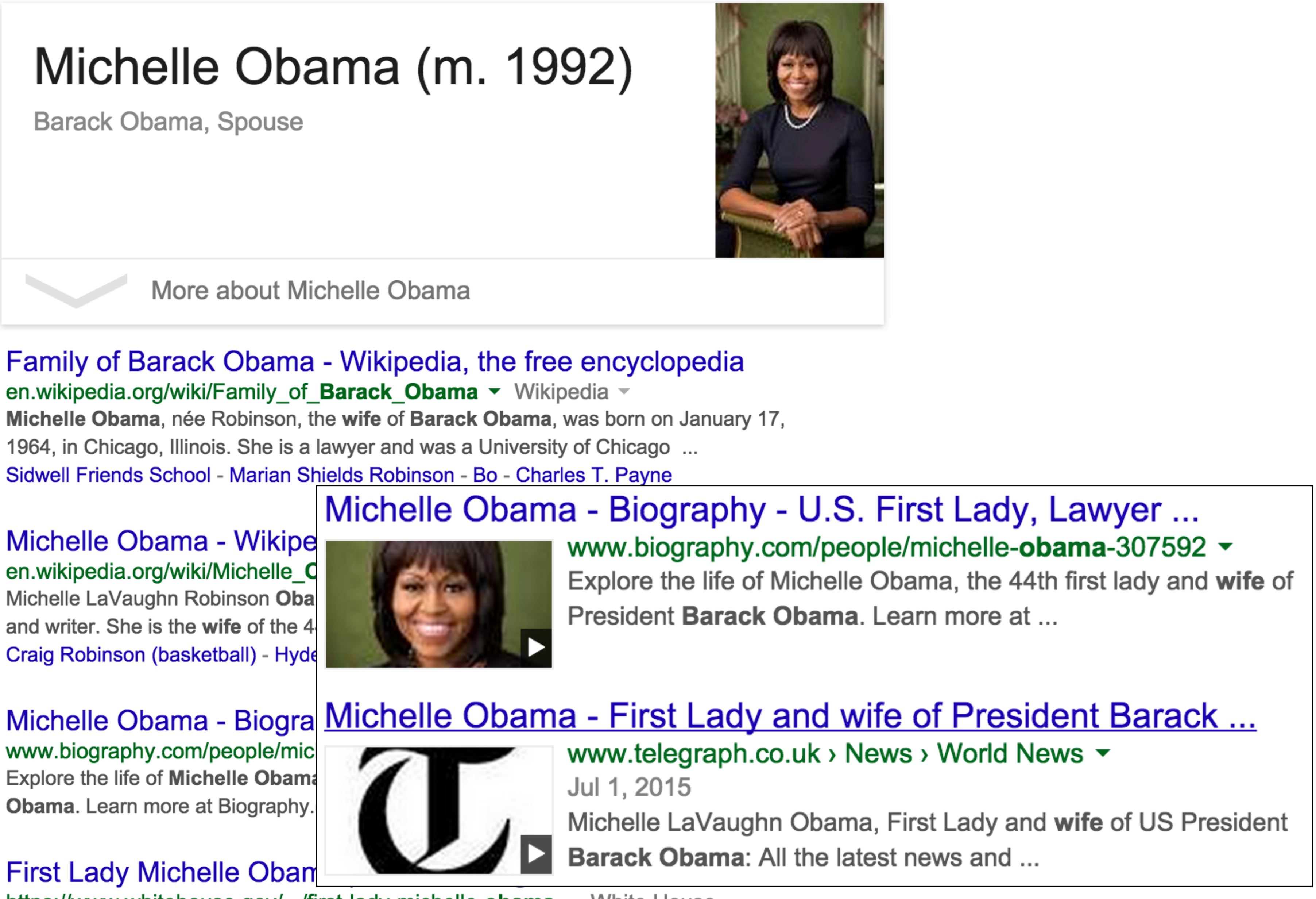
KB-LDA: Jointly Learning a Knowledge Base of Hierarchy, Relations, and Facts

Dana Movshovitz-Attias and William W Cohen

ACL, July 28, 2015

Carnegie Mellon University Computer Science Department

Who is Barack Obama's wife?



Michelle Obama (m. 1992)

Barack Obama, Spouse

More about Michelle Obama

Family of Barack Obama - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Family_of_**Barack_Obama** ▾ Wikipedia ▾
Michelle Obama, née Robinson, the **wife** of **Barack Obama**, was born on January 17, 1964, in Chicago, Illinois. She is a lawyer and was a University of Chicago ...
Sidwell Friends School - Marian Shields Robinson - Bo - Charles T. Payne

Michelle Obama - Wikipe
en.wikipedia.org/wiki/Michelle_O
Michelle LaVaughn Robinson **Oba**
and writer. She is the **wife** of the 4
Craig Robinson (basketball) - Hyde

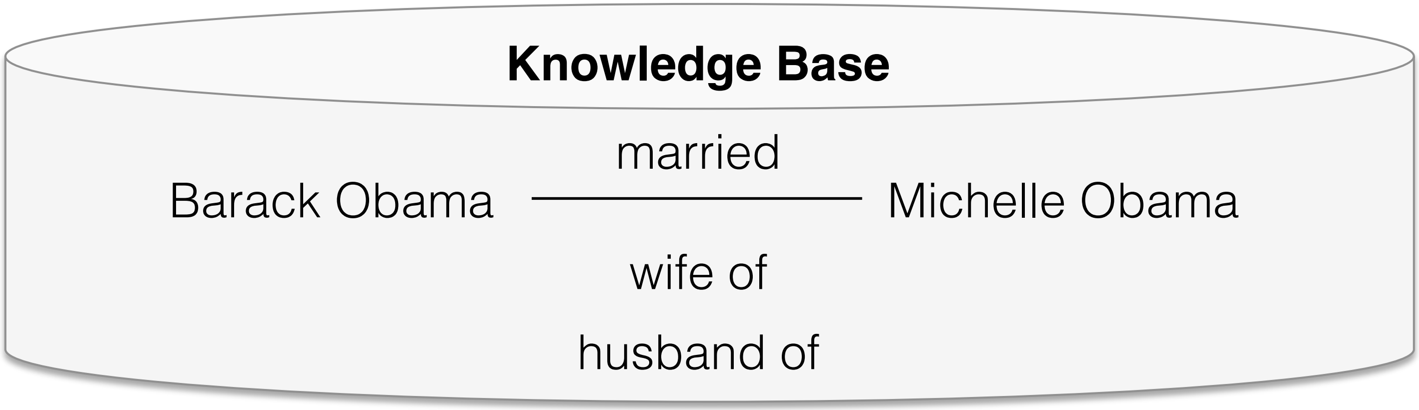
Michelle Obama - Biogra
www.biography.com/people/mic
Explore the life of **Michelle Obama**
Obama. Learn more at Biography.

First Lady Michelle Obam
https://www.whitehouse.gov/.../first-lady-michelle-obama - White House

Michelle Obama - Biography - U.S. First Lady, Lawyer ...
www.biography.com/people/michelle-**obama**-307592 ▾
Explore the life of Michelle Obama, the 44th first lady and **wife** of President **Barack Obama**. Learn more at ...

Michelle Obama - First Lady and wife of President Barack ...
www.telegraph.co.uk › News › World News ▾
Jul 1, 2015
Michelle LaVaughn Obama, First Lady and **wife** of US President **Barack Obama**: All the latest news and ...

Who is **Barack Obama**’s **wife**?



Knowledge Base

married

Barack Obama

Michelle Obama

wife of

husband of

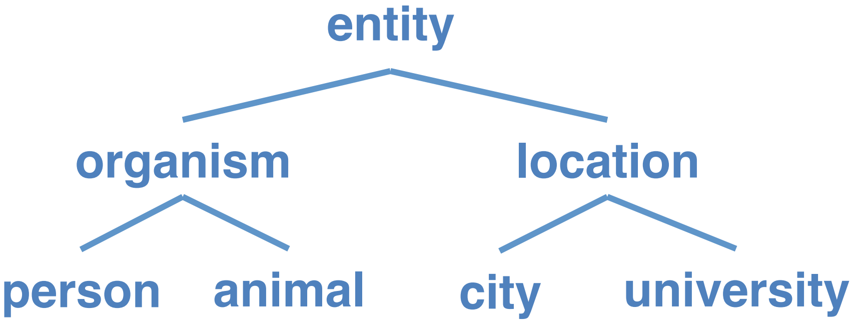
🔍 What is the **initial size** of an **ArrayList**?

🔍 What is the **Python equivalent** of a **HashMap**?

🔍 What is the **run time** of **quick sort**?

Knowledge Base (KB) Construction

Ontology-Guided Construction



entity

organism

location

person

animal

city

university

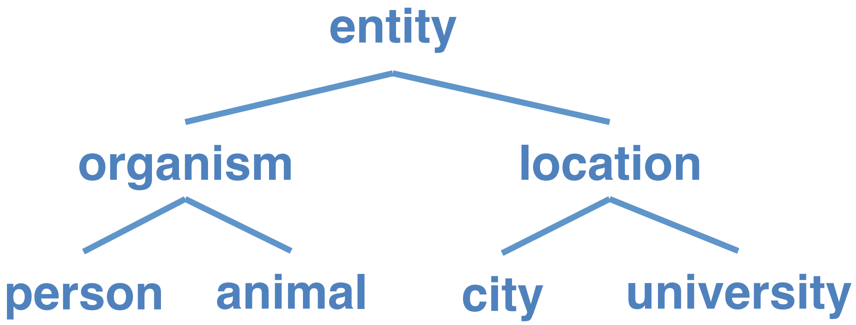
- NELL
- FreeBase
- Yago
- WordNet

Open IE

(Beijing, *is the capital of*, **China)**
(Penticton, *has very*, **warm summers)**
(Goods, *can be defined in*, **a variety of ways)**

- ReVerb
- TextRunner

Reasoning with Ontologies



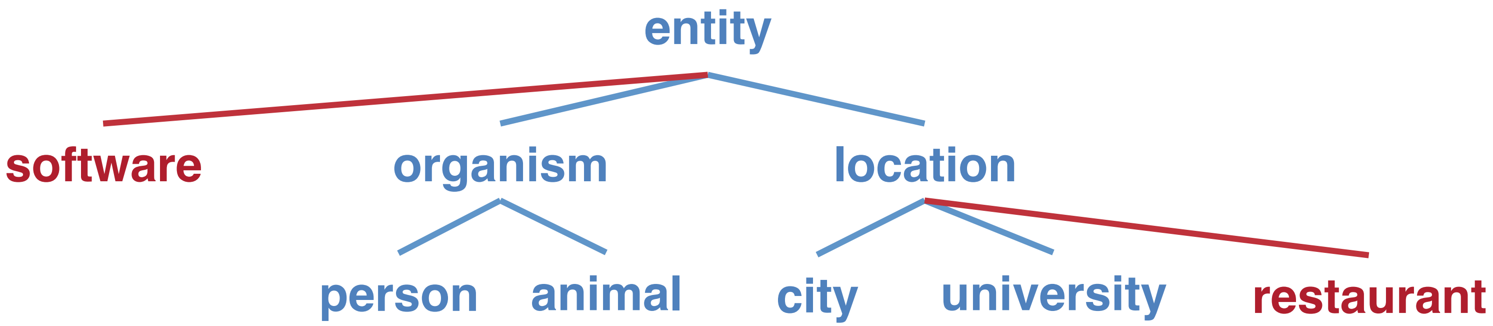
entity

organism location

person animal city university

- ✔ Ontologies give information context
- ✔ Easy to extract domain-specific information

Reasoning with Ontologies

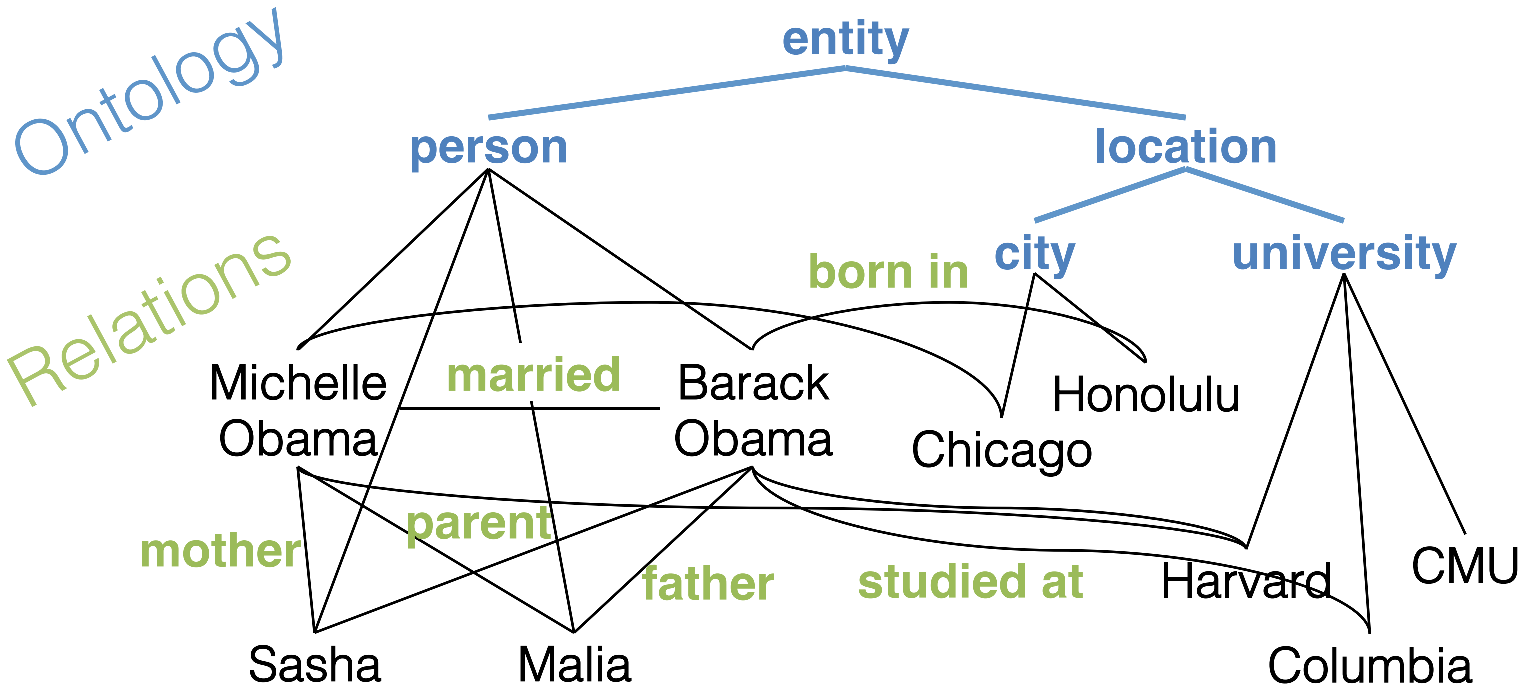


- ✓ Ontologies give information context
- ✓ Easy to extract domain-specific information

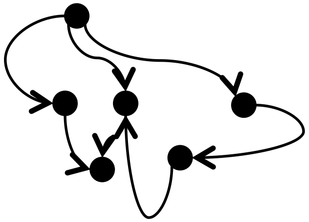
- ✗ Ontologies are
  - expensive
  - require prior knowledge
  - incomplete/outdated
- ✗ Ontology structure and facts are often drawn from different corpus statistics

Goal: Data-Driven Knowledge Base

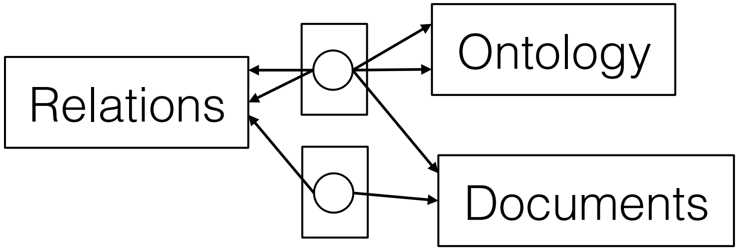
- Schema and facts are drawn from corpus and jointly optimized
- Unsupervised: learn latent corpus structure together with best-matching facts



Automatically Learning a KB with Structure and Facts



Relation Extraction

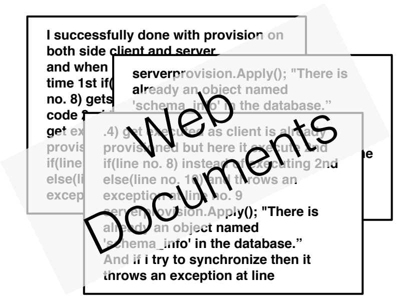


Relations

Ontology

Documents

Corpus



I successfully done with provision on both side client and server and when time 1st if(no. 8) gets code get ex provis if(line else(li excep

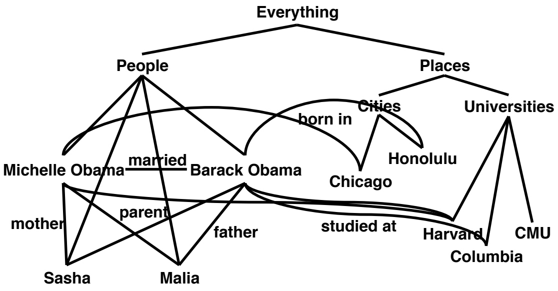
serverprovision.Apply(); "There is already an object named 'schema_info' in the database."

.4) get executed as client is already provisioned but here it execute 2nd if(line no. 8) instead of executing 2nd else(line no. 10) and throws an exception at line no. 9 serverprovision.Apply(); "There is already an object named 'schema_info' in the database." And if i try to synchronize then it throws an exception at line

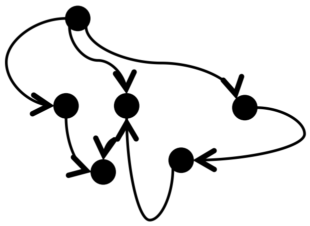
Collection of Documents from StackOverflow

KB-LDA

KB Evaluation

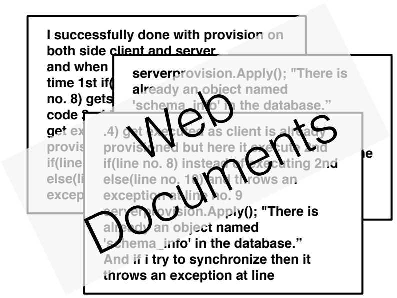


Automatically Learning a KB with Structure and Facts



Relation Extraction

Corpus

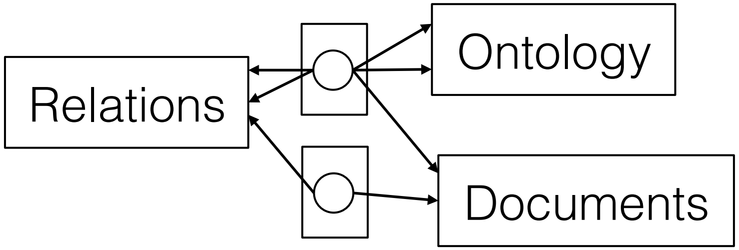


I successfully done with provision on both side client and server and when time 1st if(no. 8) gets code f get ex provis if(line else(li excep

serverprovision.Apply(); "There is already an object named 'schema_info' in the database."

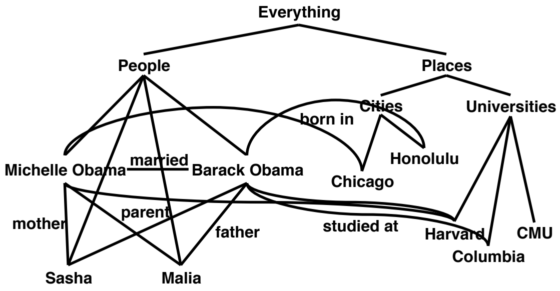
.4) get executed as client is already provisioned but here it execute 2nd if(line no. 8) instead of executing 2nd else(line no. 10) and throws an exception at line no. 9 serverprovision.Apply(); "There is already an object named 'schema_info' in the database." And if i try to synchronize then it throws an exception at line

Collection of Documents from StackOverflow



KB-LDA

KB Evaluation

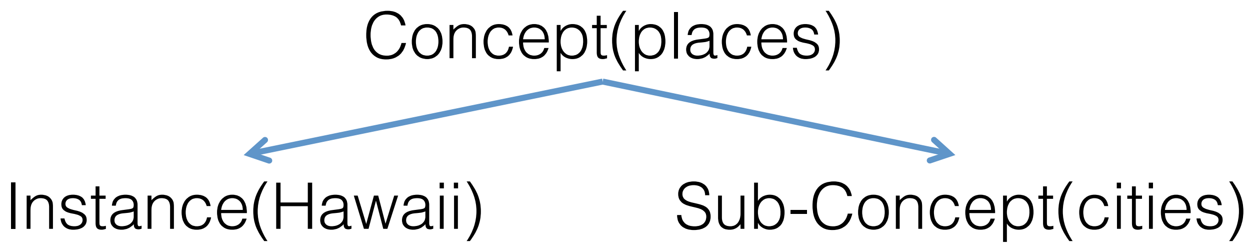


Pattern-based Relation Extraction

Ontology

“**places** such as **Hawaii**”

“**places** including **cities**, towns and villages”



Hearst patterns	Hypernym-hyponyms from StackOverflow
Y such as **X**	**Languages** such as **Java**
X is a **Y**	**IDEs** including **Eclipse** and **NetBeans**
Y including **X**	**Integer** and other **data types**

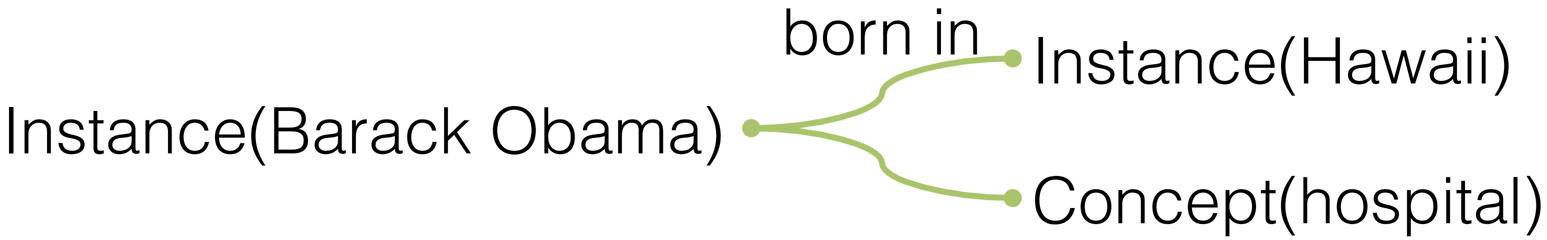
Pattern-based Relation Extraction

Relations

Subject-Verb-Object patterns

“**Barack Obama** was born in **Hawaii**”

“**Barack Obama** was born in **a hospital**”



born in

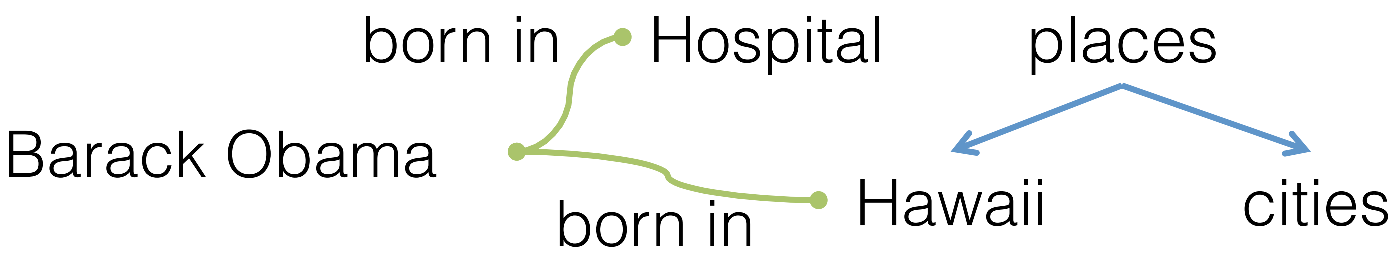
Instance(Barack Obama)

Instance(Hawaii)

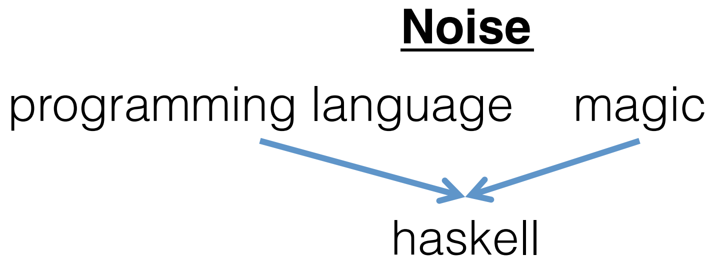
Concept(hospital)

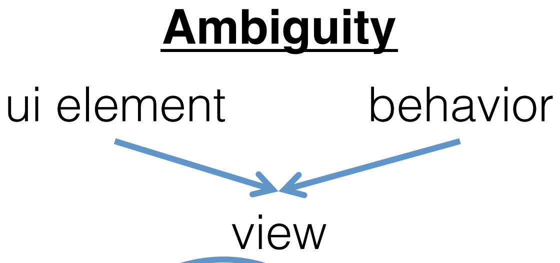
SVOs from StackOverflow
Class implements **interface**
Constructor throws **exception**
Output fills **buffer**

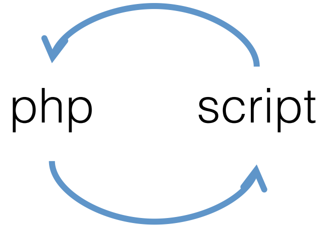
Open Structure of Extracted Relations



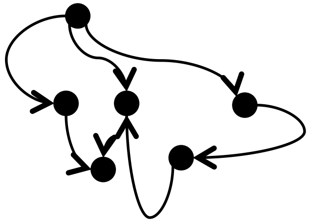
- No distinction between concepts and instances
- Parsing errors: “**class** not found **exception**”



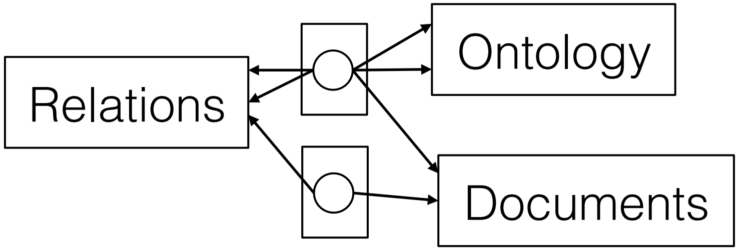


- Structure includes cycles: 

Automatically Learning a KB with Structure and Facts

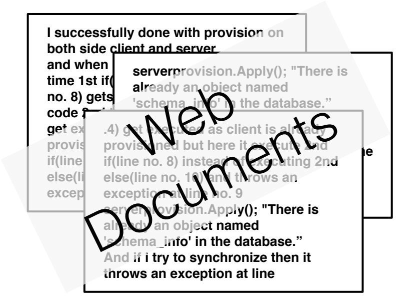


Relation Extraction



KB-LDA

Corpus



Web Documents

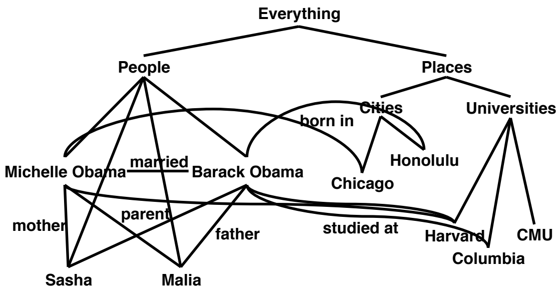
I successfully done with provision on
both side client and server
and when
time 1st if(
no. 8) gets
code 2
get ex
provis
if(line
else(li
excep

serverprovision.Apply(); "There is
already an object named
'schema_info' in the database."

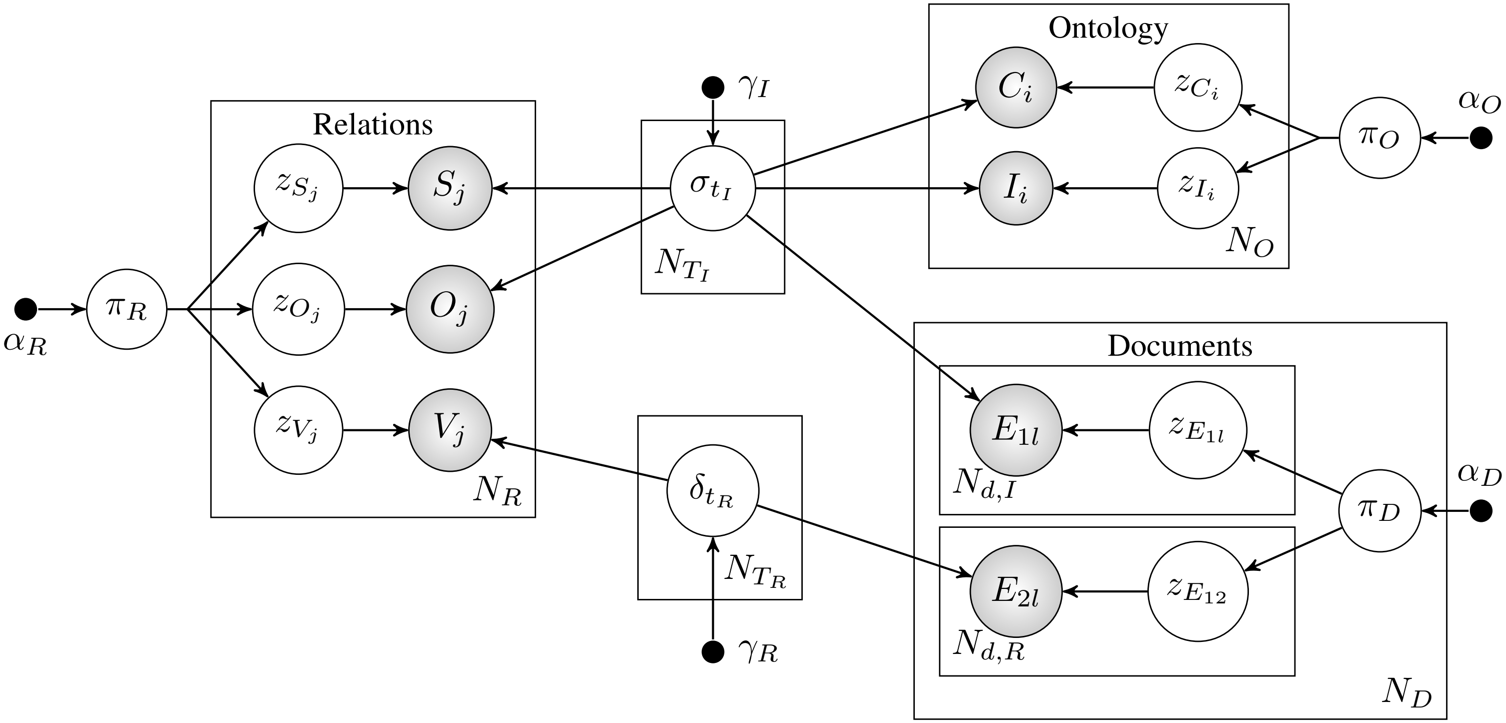
.4) get executed as client is already
provisioned but here it execute 2nd
if(line no. 8) instead of executing 2nd
else(line no. 10) and throws an
exception at line no. 9
serverprovision.Apply(); "There is
already an object named
'schema_info' in the database."
And if i try to synchronize then it
throws an exception at line

Collection of Documents from StackOverflow

KB Evaluation

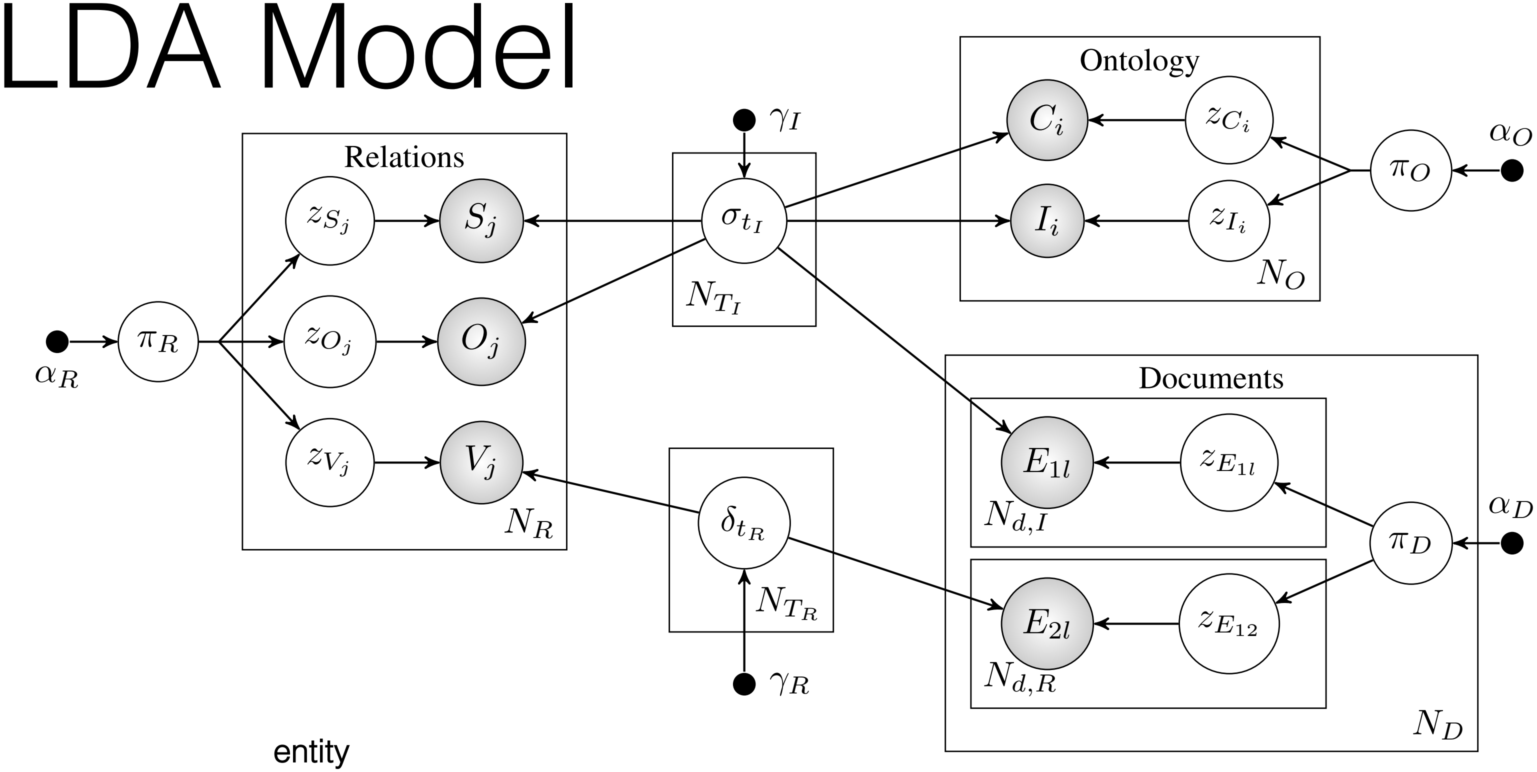


KB-LDA Model

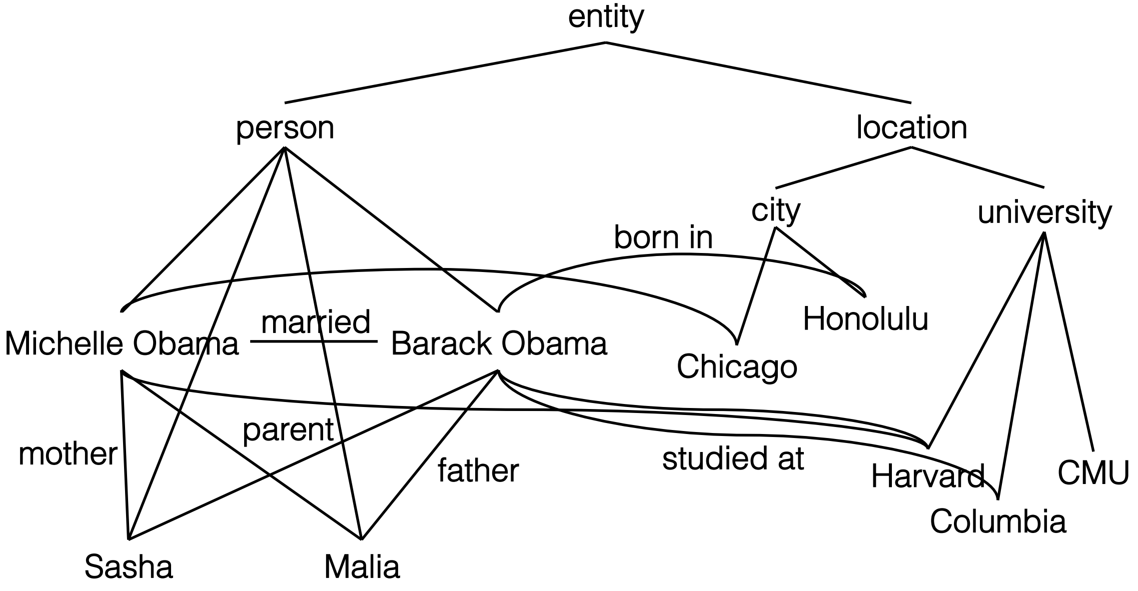


Relations
z_{S_j}
S_j
z_{O_j}
O_j
z_{V_j}
V_j
N_R
α_R
π_R
γ_I
σ_{t_I}
N_{T_I}
δ_{t_R}
N_{T_R}
γ_R
Ontology
C_i
z_{C_i}
I_i
z_{I_i}
N_O
π_O
α_O
Documents
E_{1l}
$z_{E_{1l}}$
$N_{d,I}$
E_{2l}
$z_{E_{12}}$
$N_{d,R}$
N_D
π_D
α_D

KB-LDA Model

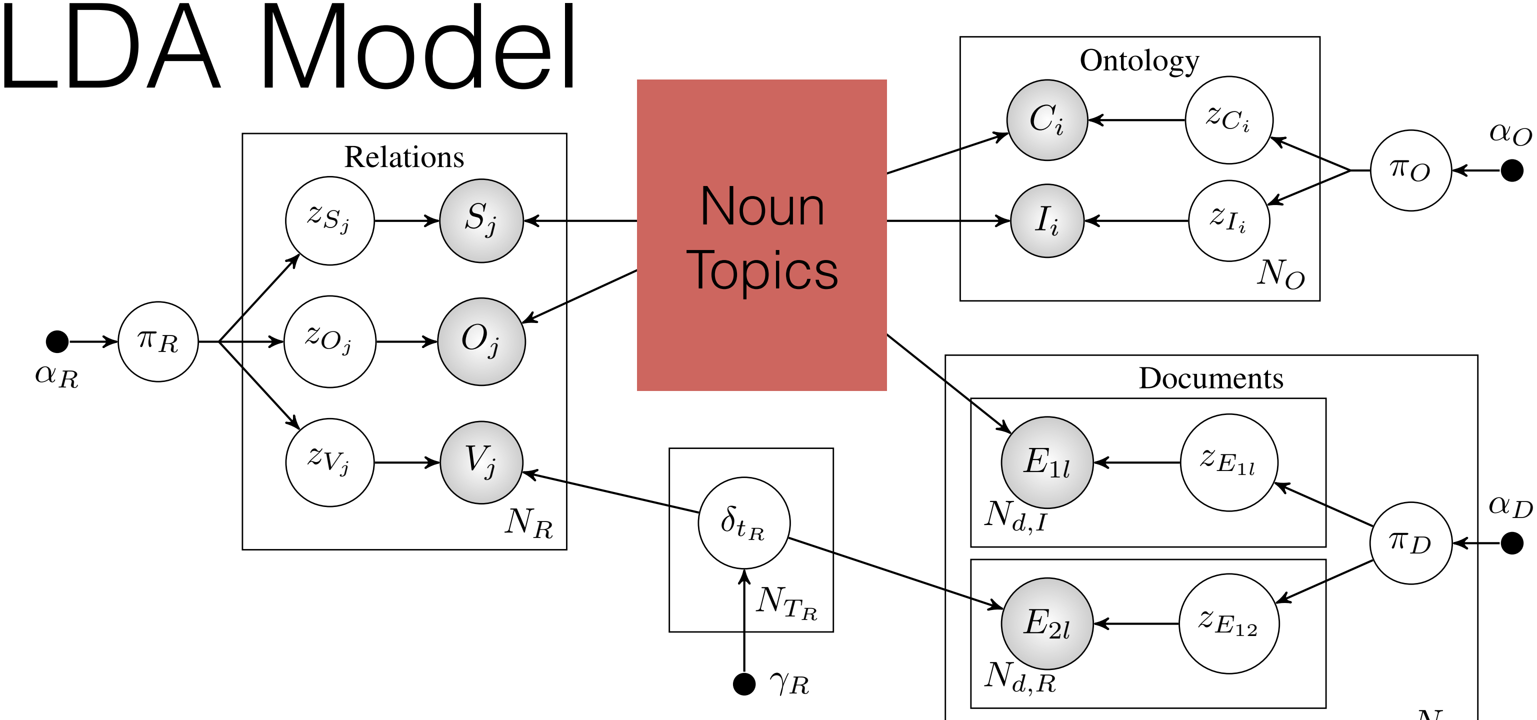


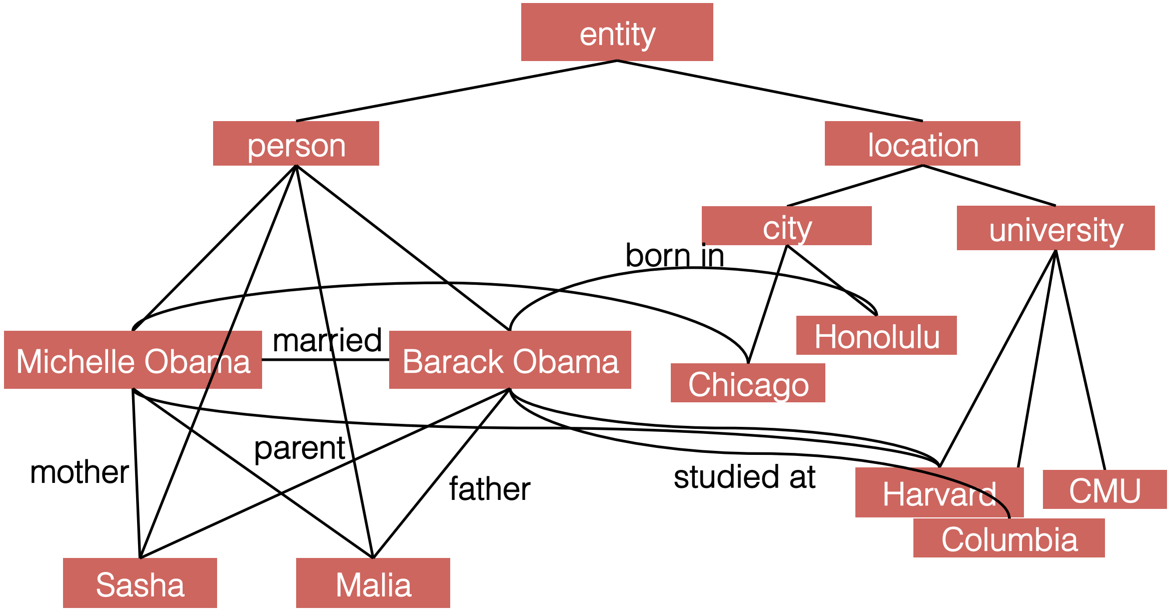
Relations
z_{S_j}
S_j
z_{O_j}
O_j
z_{V_j}
V_j
N_R
α_R
π_R
γ_I
σ_{t_I}
N_{T_I}
δ_{t_R}
N_{T_R}
γ_R
Ontology
C_i
z_{C_i}
I_i
z_{I_i}
N_O
π_O
α_O
Documents
E_{1l}
$z_{E_{1l}}$
$N_{d,I}$
E_{2l}
$z_{E_{12}}$
$N_{d,R}$
N_D
π_D
α_D



entity
person
location
city
university
born in
Michelle Obama
married
Barack Obama
Chicago
Honolulu
mother
parent
father
studied at
Harvard
CMU
Columbia
Sasha
Malia

KB-LDA Model





Michelle Obama

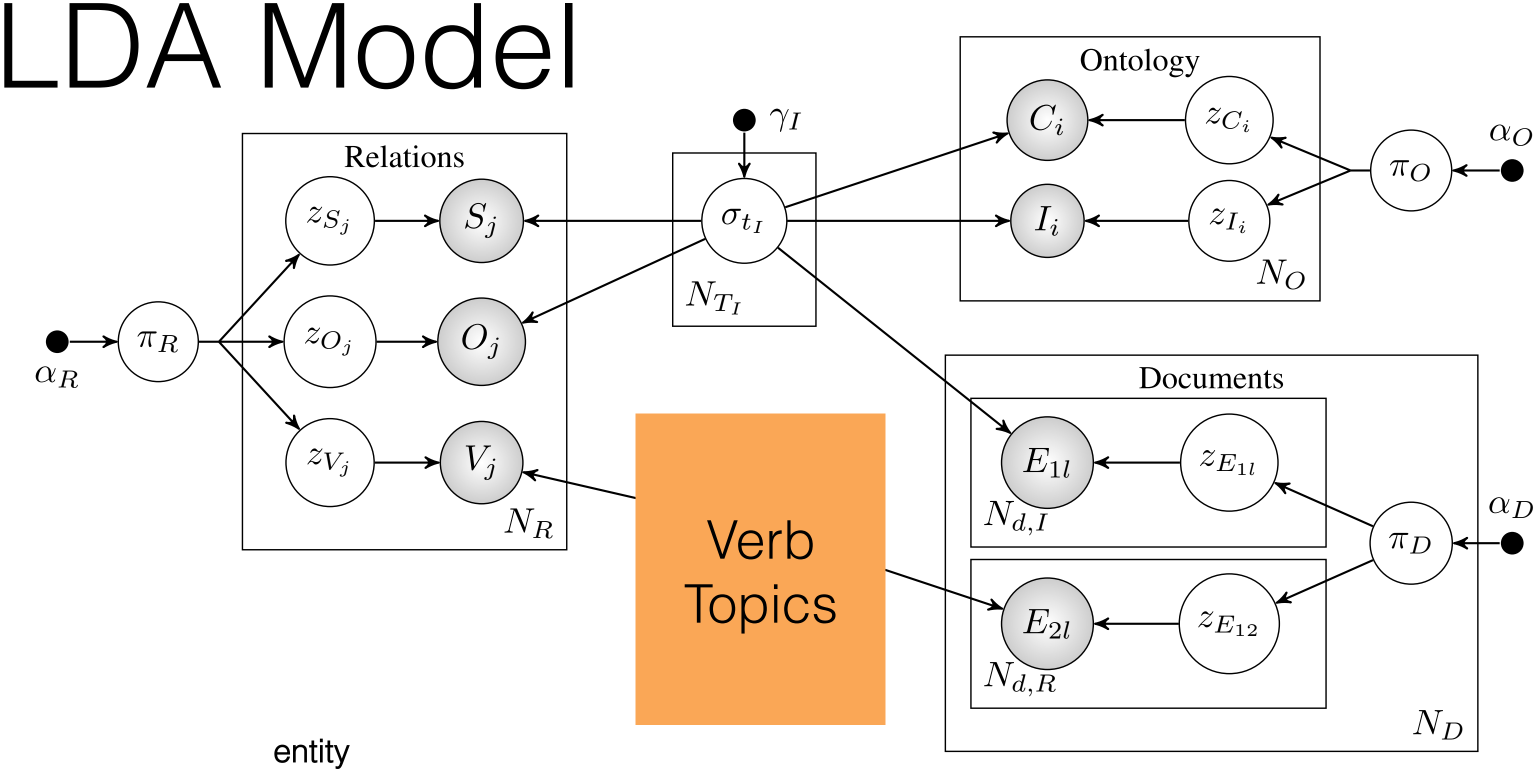
Barack Obama

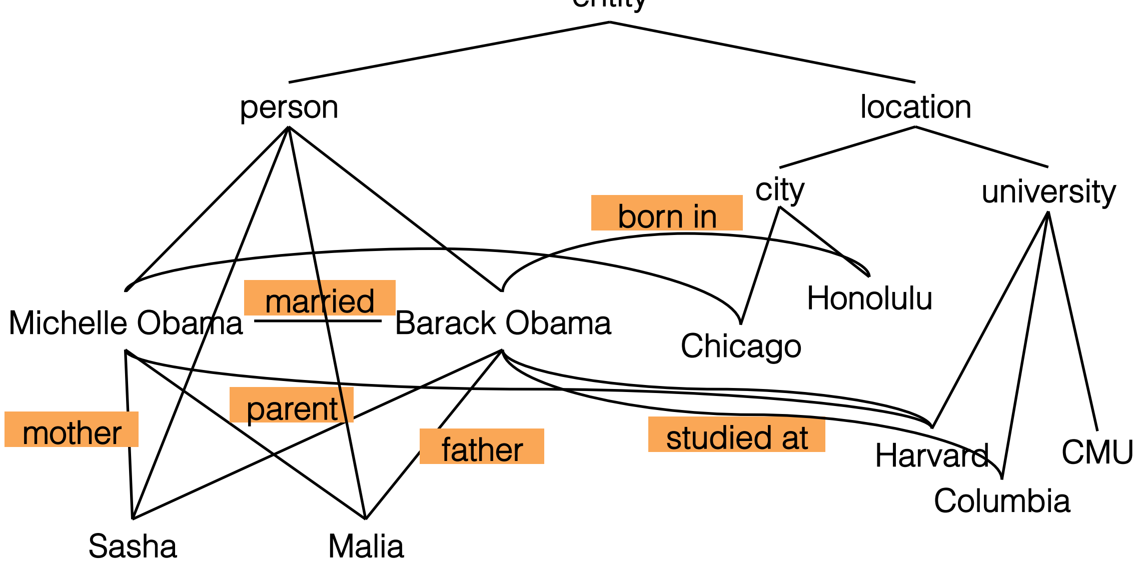
Malia

Sasha

. . .

KB-LDA Model





married to

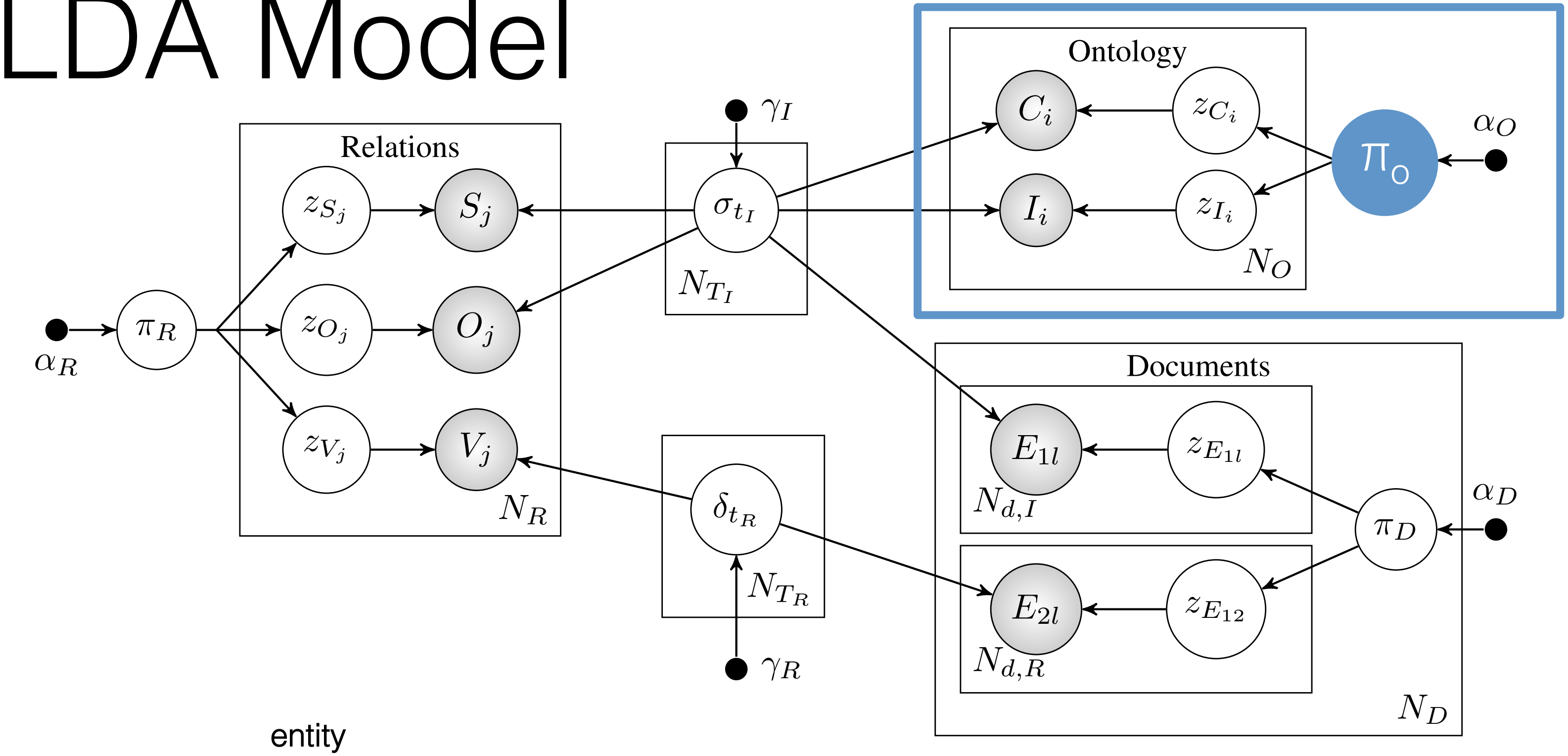
is husband of

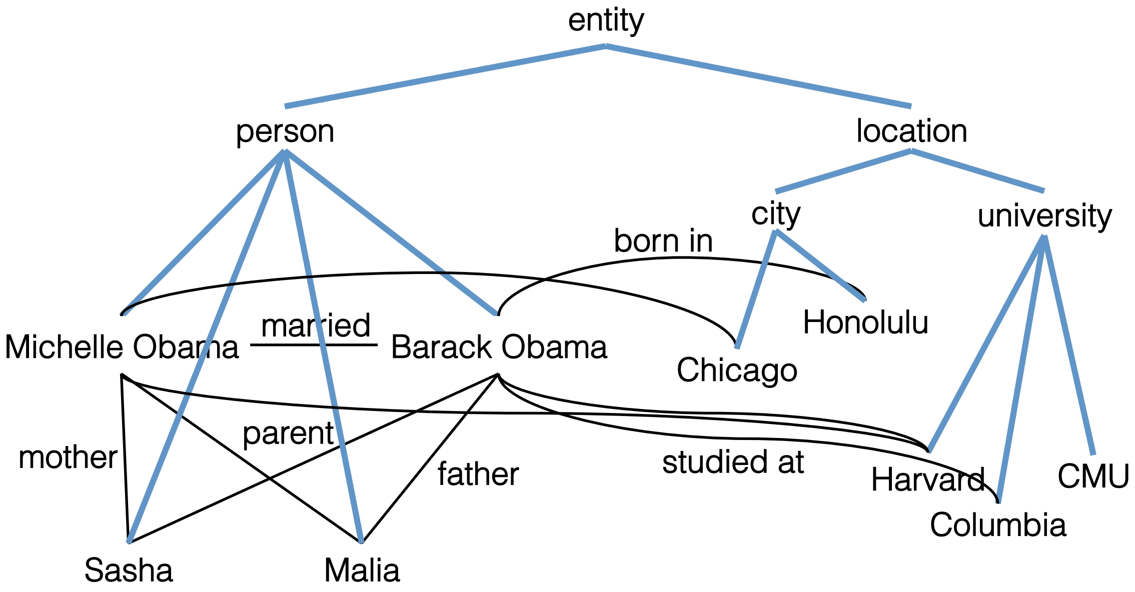
is wife of

. . .

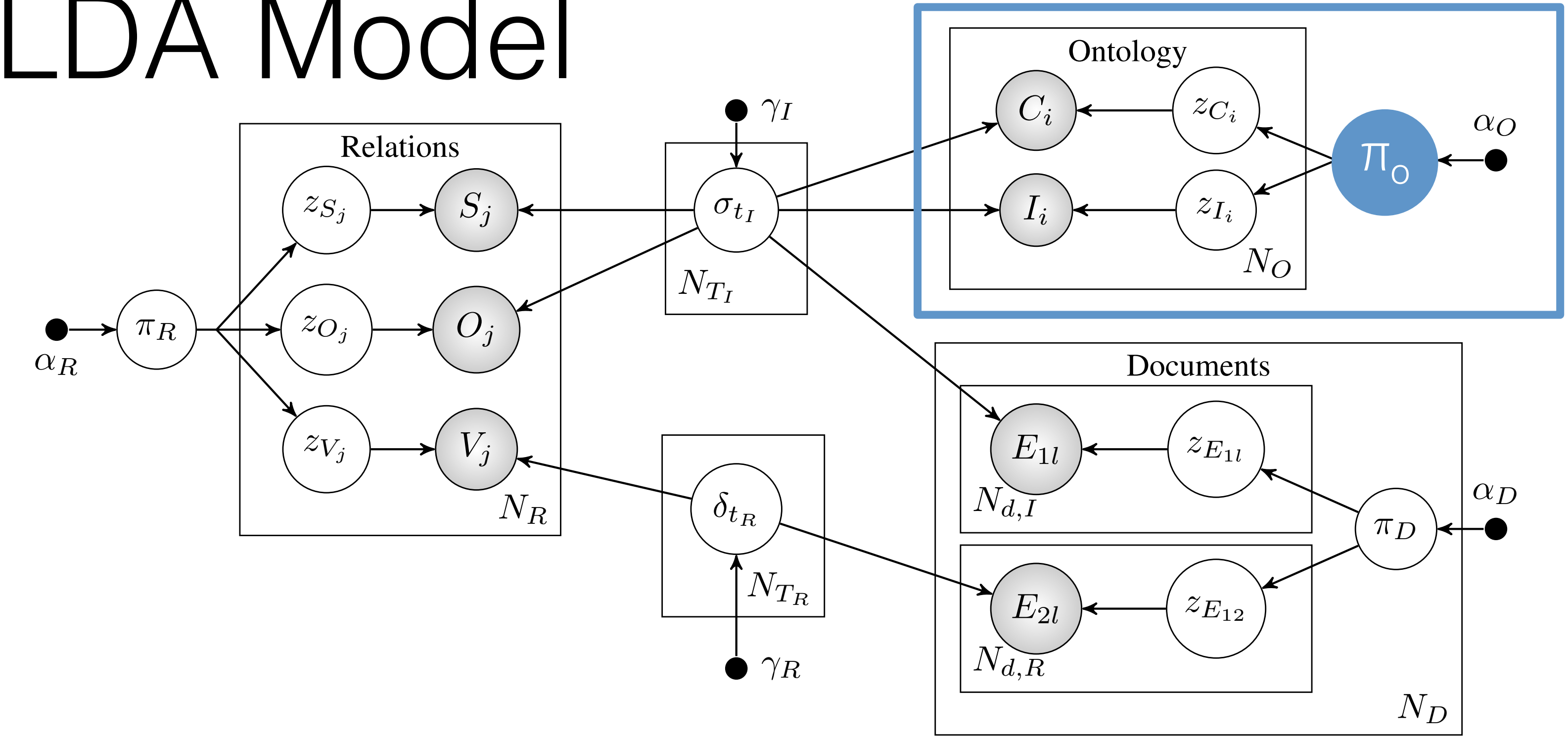
KB-LDA Model

Ontology





KB-LDA Model

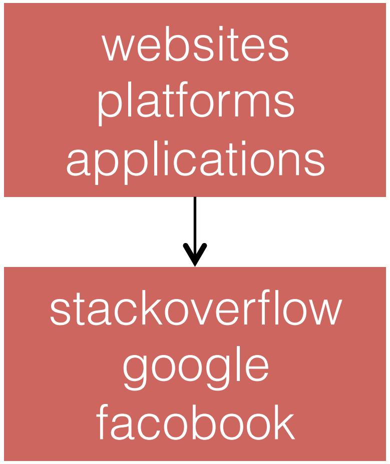


Ontology

Extracted hypernym-hyponym relations:

websites → google

platforms → stackoverflow



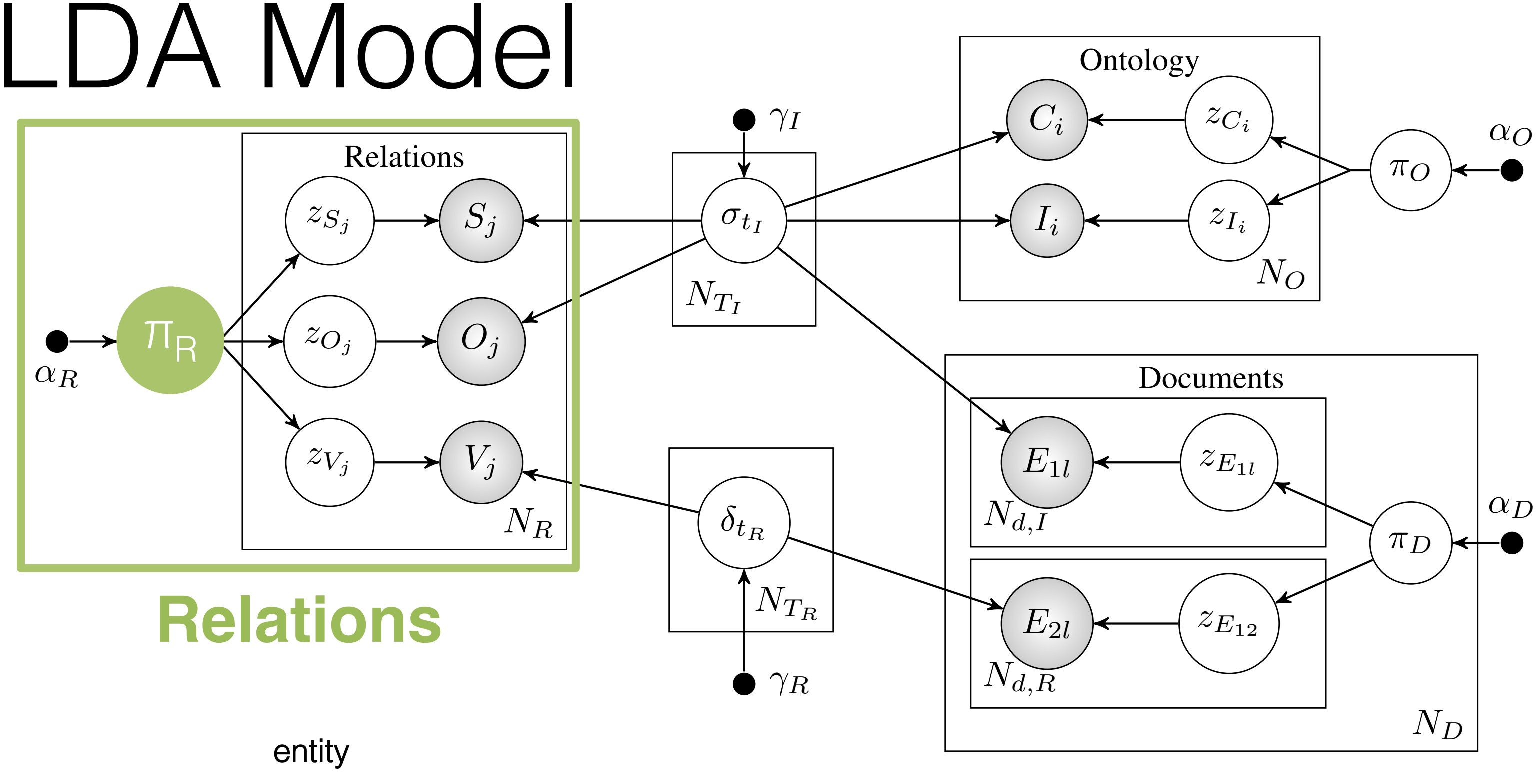
websites
platforms
applications

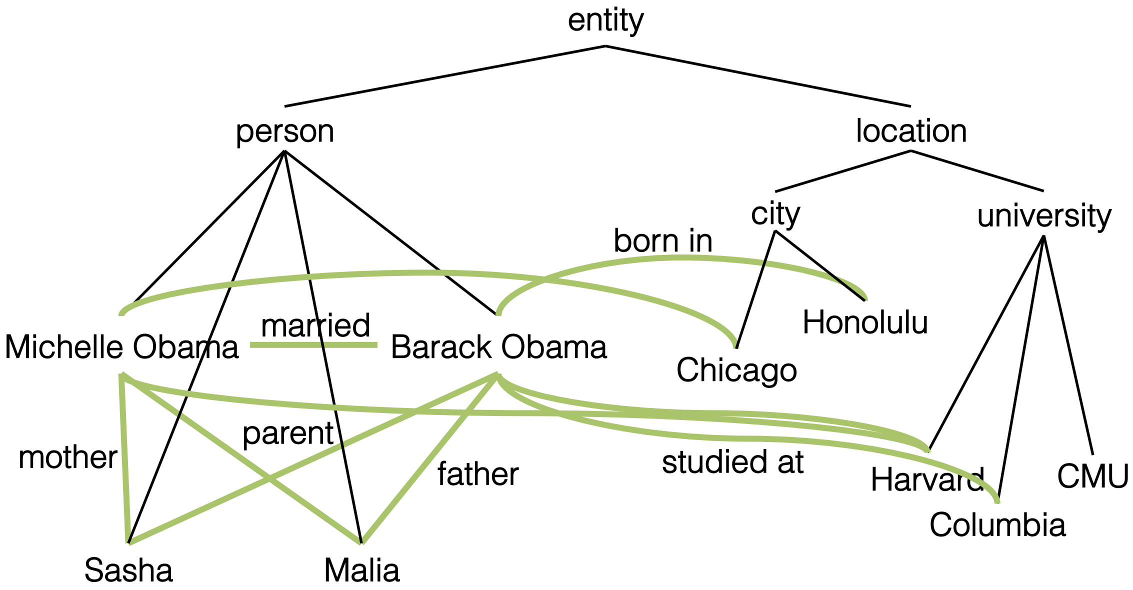
Noun Topic 1

stackoverflow
google
facobook

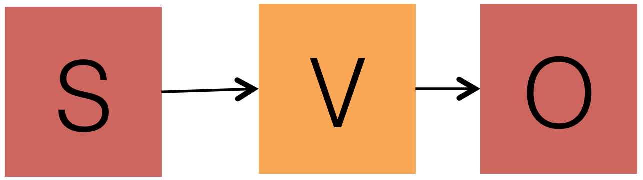
Noun Topic 2

KB-LDA Model

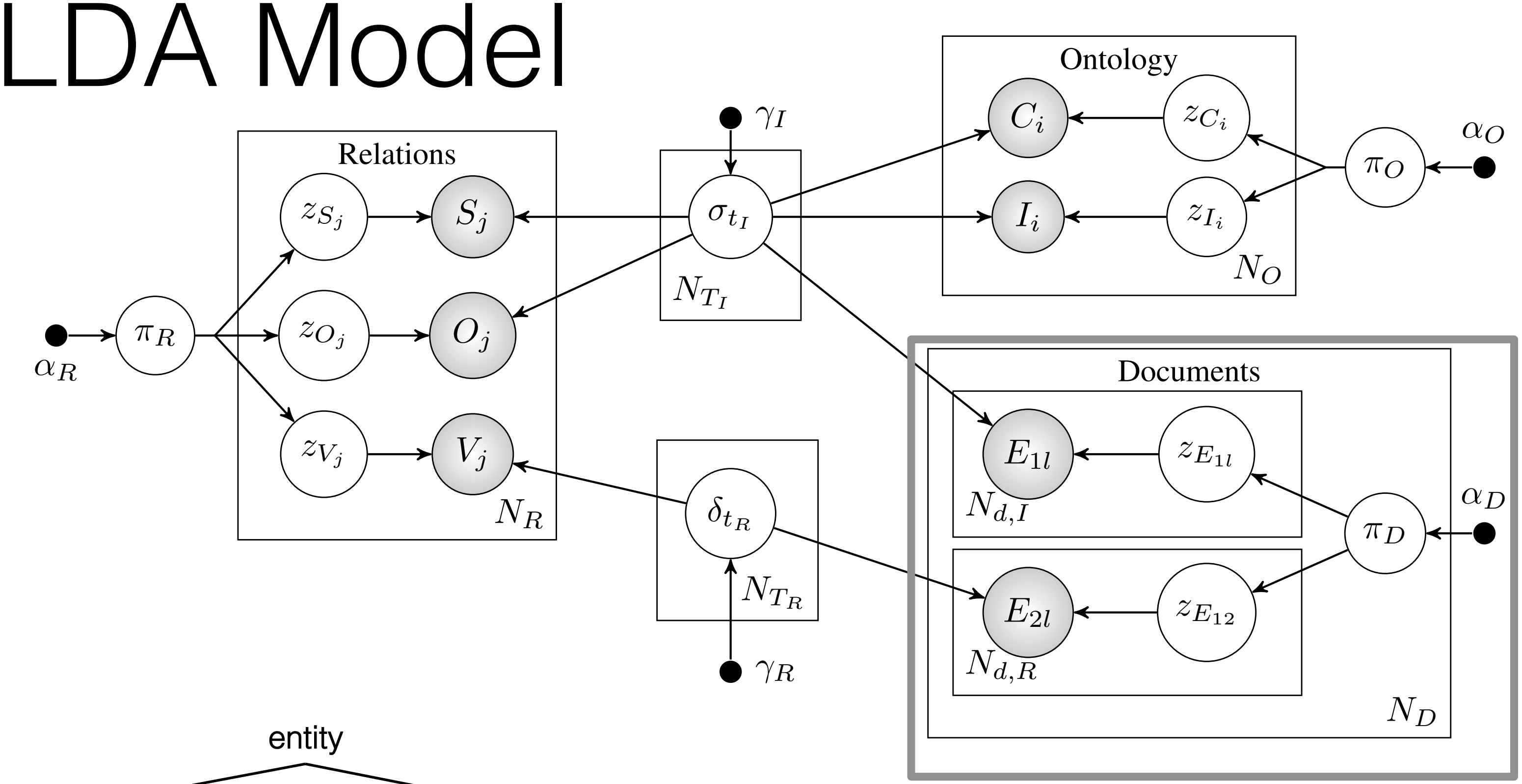




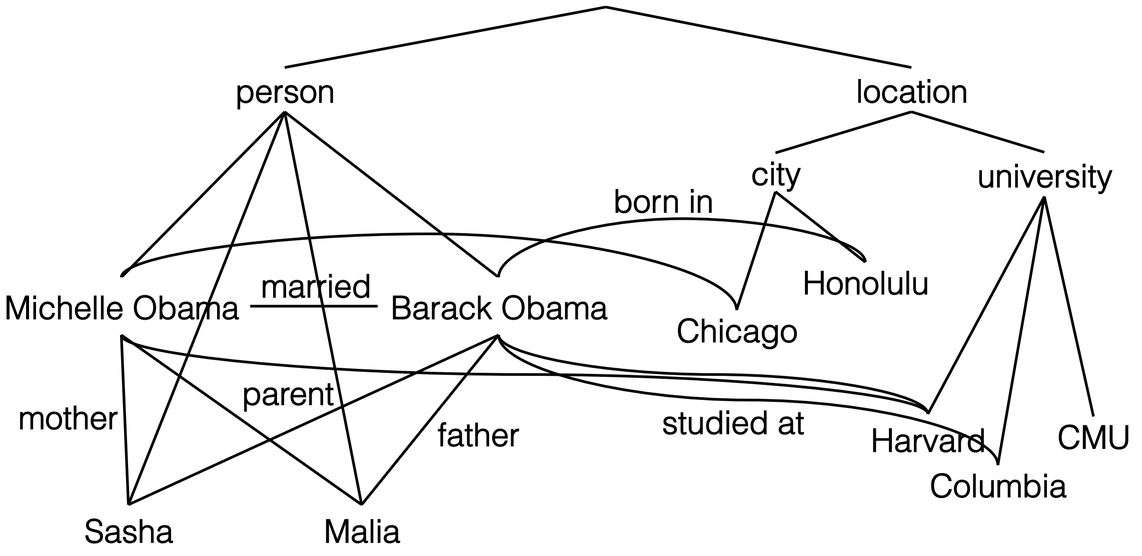
Extracted SVO relation:
Michelle, *wife of*, Barack



KB-LDA Model

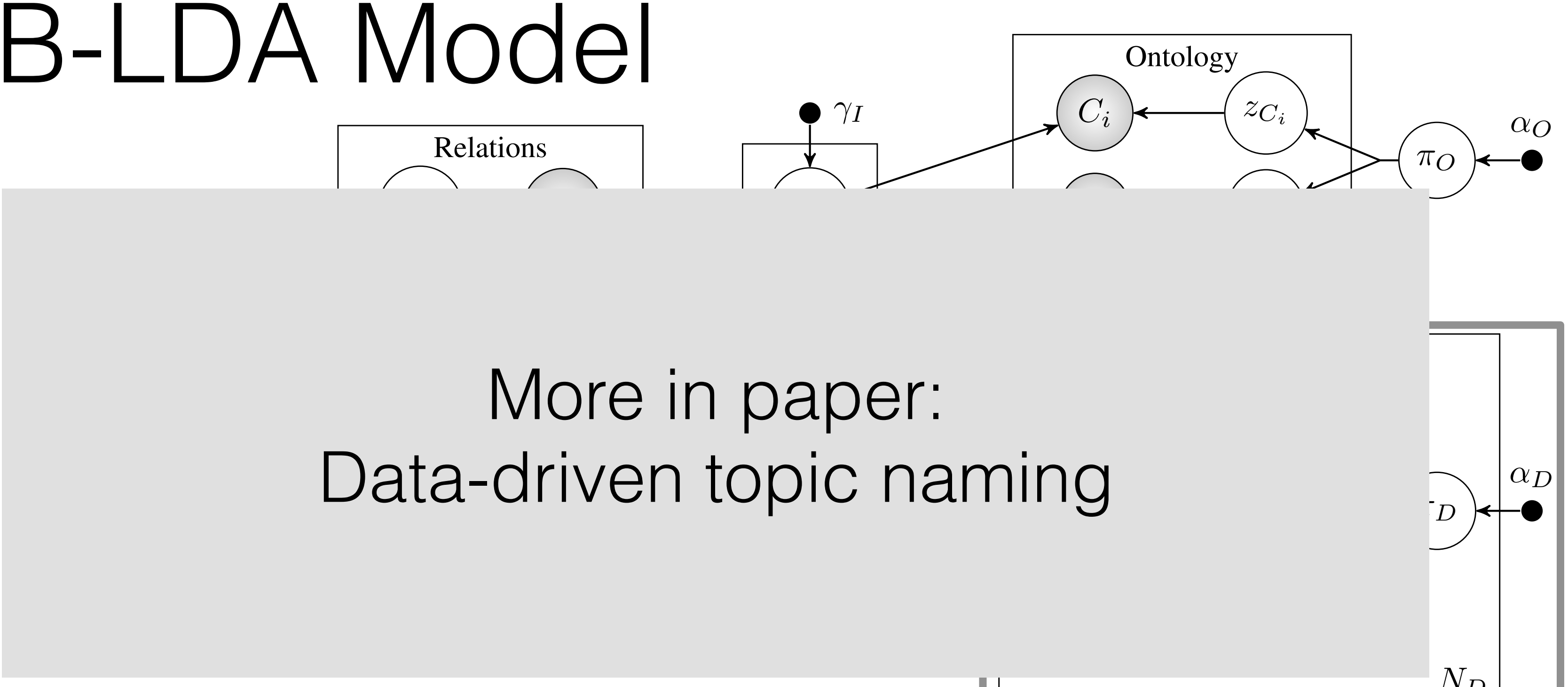


Documents

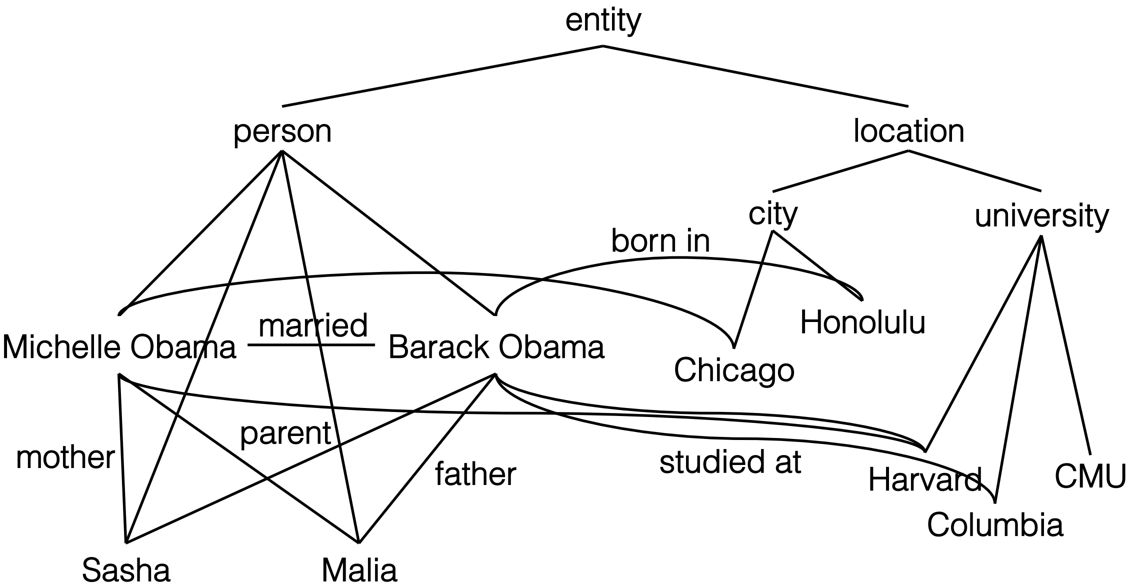


Downloads on websites sometimes have an **MD5** checksum, **allowing people** to confirm the integrity of the file. I have heard this is to allow not only corrupted files to be instantly identified before they cause a problem but also for for any **malicious changes** to be easily **detected**.

KB-LDA Model



More in paper:
Data-driven topic naming



Documents

Downloads on websites sometimes have an **MD5** checksum, **allowing people** to confirm the integrity of the file. I have heard this is to allow not only corrupted files to be instantly identified before they cause a problem but also for for any **malicious changes** to be easily **detected**.

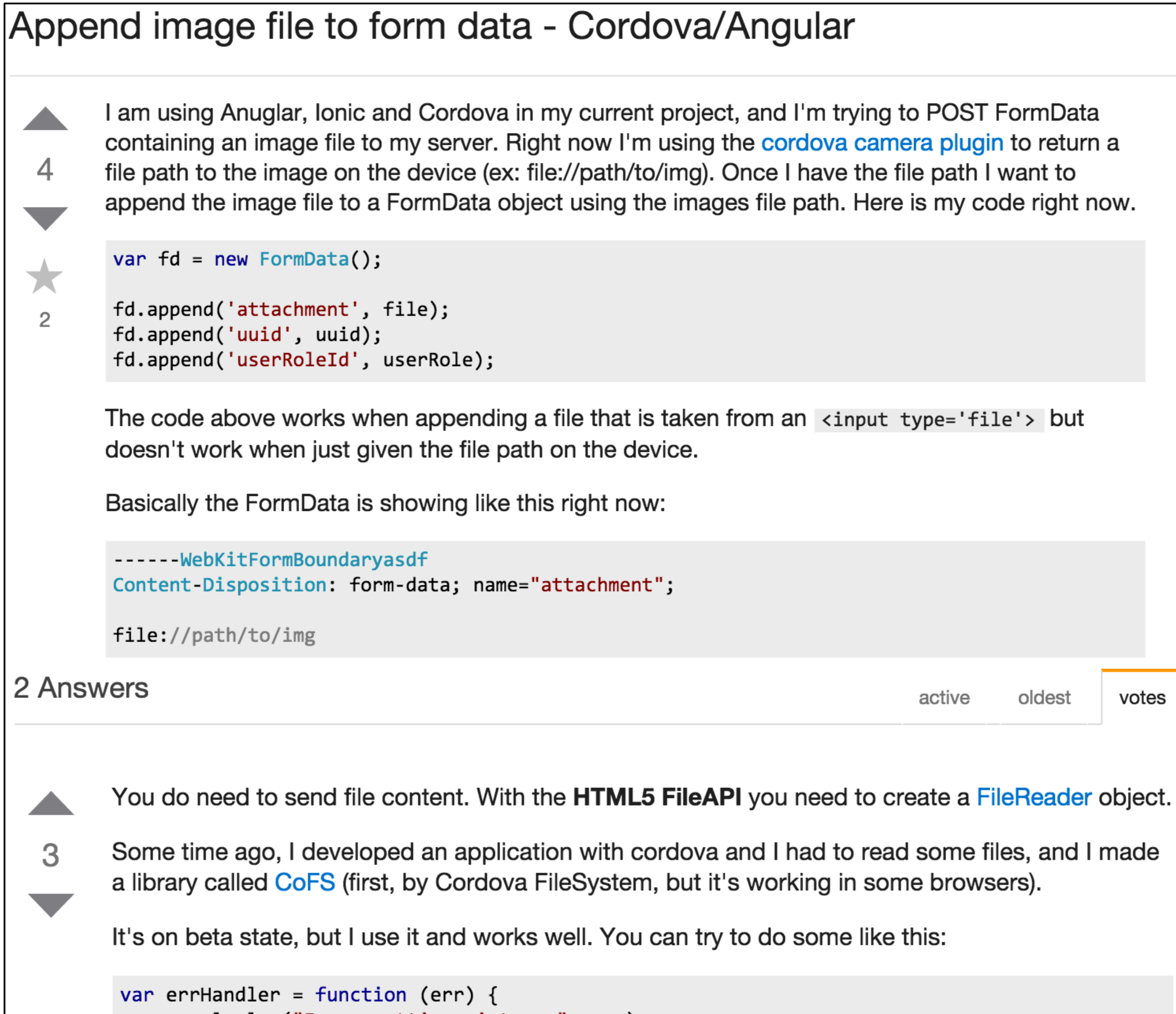
Training

Corpus

5.5M documents from StackOverflow

Topics

50



Append image file to form data - Cordova/Angular

I am using Anuglar, Ionic and Cordova in my current project, and I'm trying to POST FormData containing an image file to my server. Right now I'm using the cordova camera plugin to return a file path to the image on the device (ex: file://path/to/img). Once I have the file path I want to append the image file to a FormData object using the images file path. Here is my code right now.

```
var fd = new FormData();

fd.append('attachment', file);
fd.append('uuid', uuid);
fd.append('userRoleId', userRole);
```

The code above works when appending a file that is taken from an `<input type='file'>` but doesn't work when just given the file path on the device.

Basically the FormData is showing like this right now:

```
------WebKitFormBoundaryasdf
Content-Disposition: form-data; name="attachment";

file://path/to/img
```

2 Answers

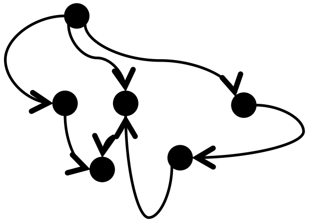
You do need to send file content. With the **HTML5 FileAPI** you need to create a FileReader object.

Some time ago, I developed an application with cordova and I had to read some files, and I made a library called CoFS (first, by Cordova FileSystem, but it's working in some browsers).

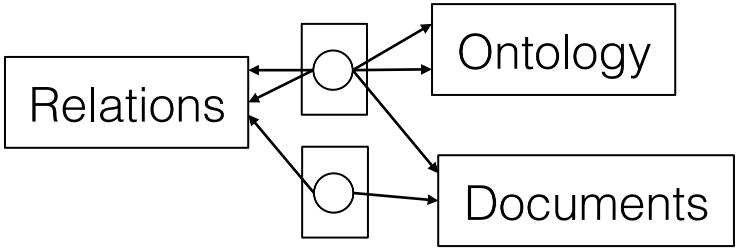
It's on beta state, but I use it and works well. You can try to do some like this:

```
var errHandler = function (err) {
```

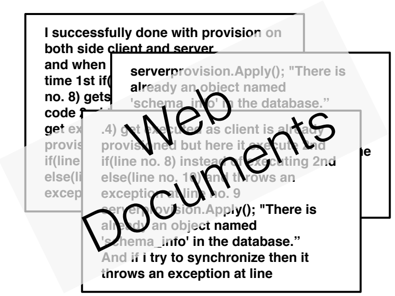
Automatically Learning a KB with Structure and Facts



Relation Extraction



Corpus



I successfully done with provision on both side client and server and when time 1st if(no. 8) gets code get ex provis if(line else(li excep

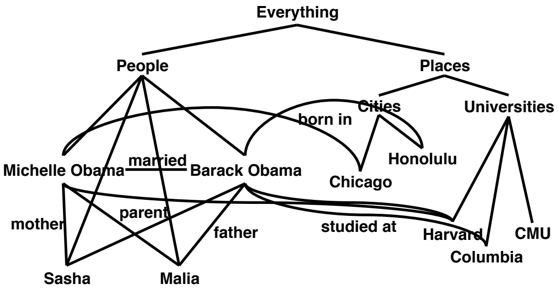
serverprovision.Apply(); "There is already an object named 'schema_info' in the database."

.4) get executed as client is already provisioned but here it execute 2nd if(line no. 8) instead of executing 2nd else(line no. 10) and throws an exception at line no. 9 serverprovision.Apply(); "There is already an object named 'schema_info' in the database." And if i try to synchronize then it throws an exception at line

Collection of Documents from StackOverflow

KB-LDA

KB Evaluation



Top Tokens of Learned Noun Topics

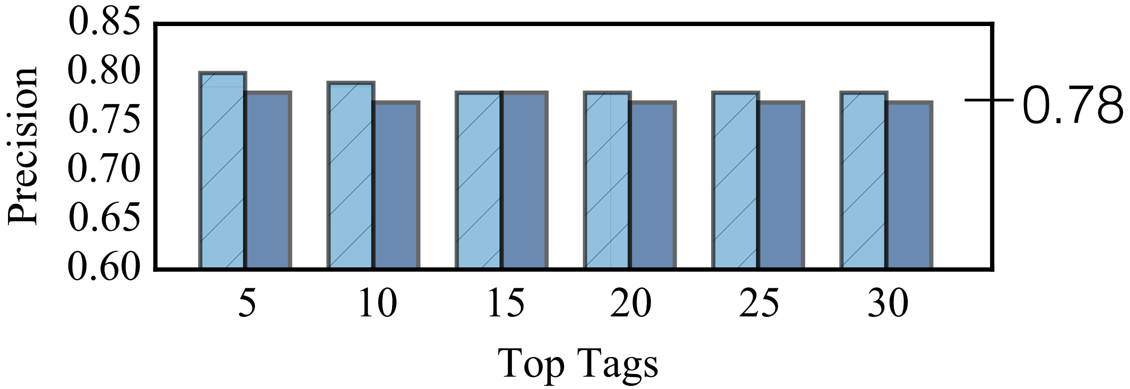
database	**web browsers**	**programming language**	**user information**
table	ie	java	name
query	firefox	python	images
database	chrome	javascript	id
sql	safari	lists	number
column	explorer	ruby	text
data	ie7	c +	password
tables	ie6	perl	address
mysql	ie8	haskell	strings
index	ie9	jquery	files
columns	internet explorer	scala	string

M-Turk Evaluation of Noun Topics

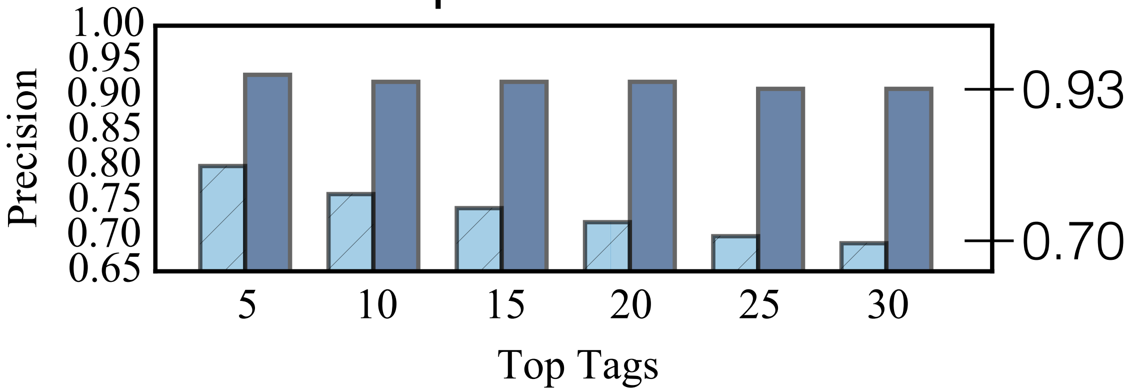
“Which words are ***not*** related to **programming languages**?”

java
python
javascript
firefox
ruby
perl

1. Word Intrusion



2. Group Precision

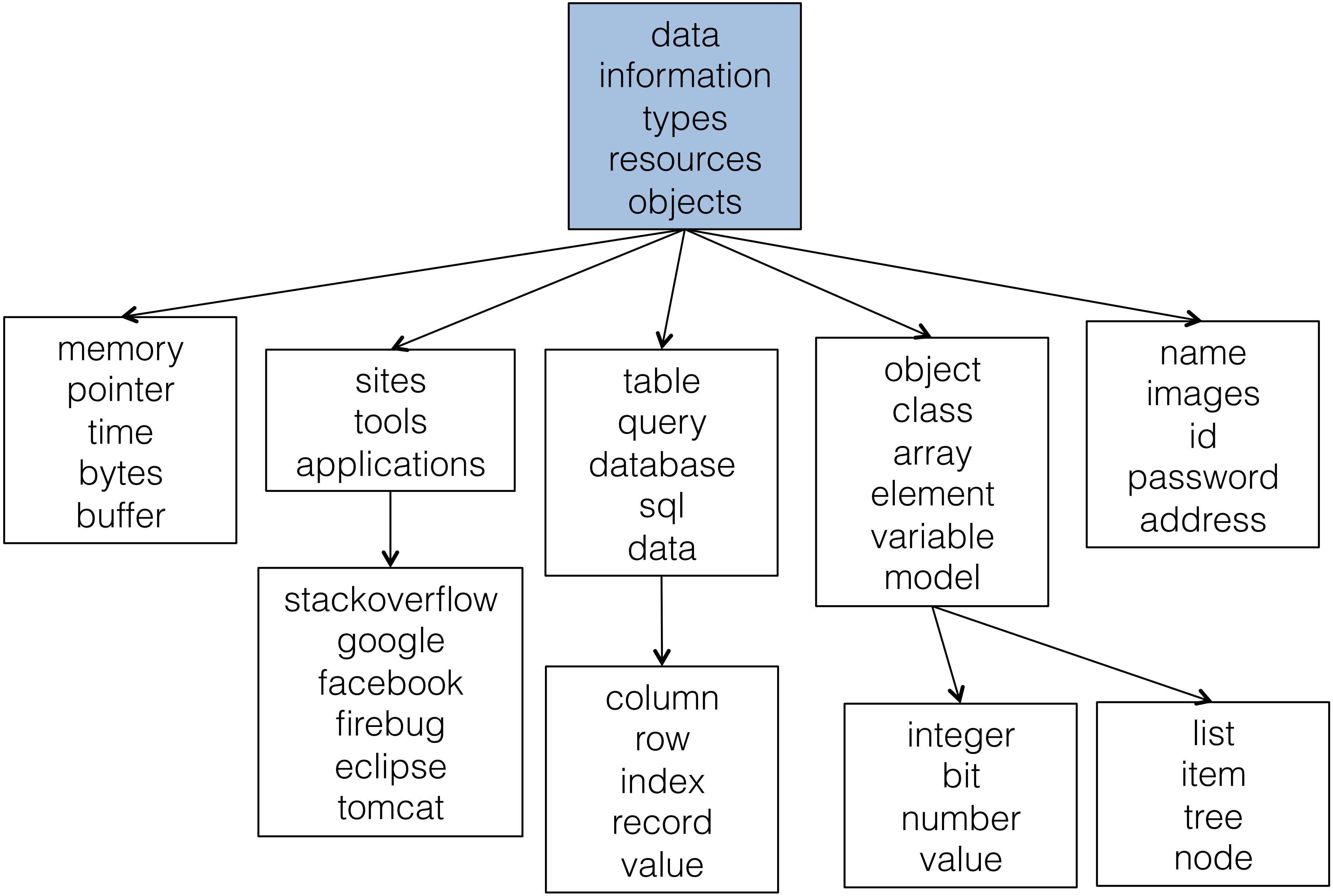


Relations

Users ---------->	interact with ------->	**UI element**
user	clicks	button
people	selects	form
customer	submits	link
client	hits	item
player	moves	file

“Is this a reasonable Software relationship?”

	KB-LDA	**Random**
Avg(Experts)	0.7	0.13
1 Worker	0.9	0.69
2 Workers	0.63	0.22
3 Workers	0.15	0.05



data
information
types
resources
objects

memory
pointer
time
bytes
buffer

sites
tools
applications

stackoverflow
google
facebook
firebug
eclipse
tomcat

table
query
database
sql
data

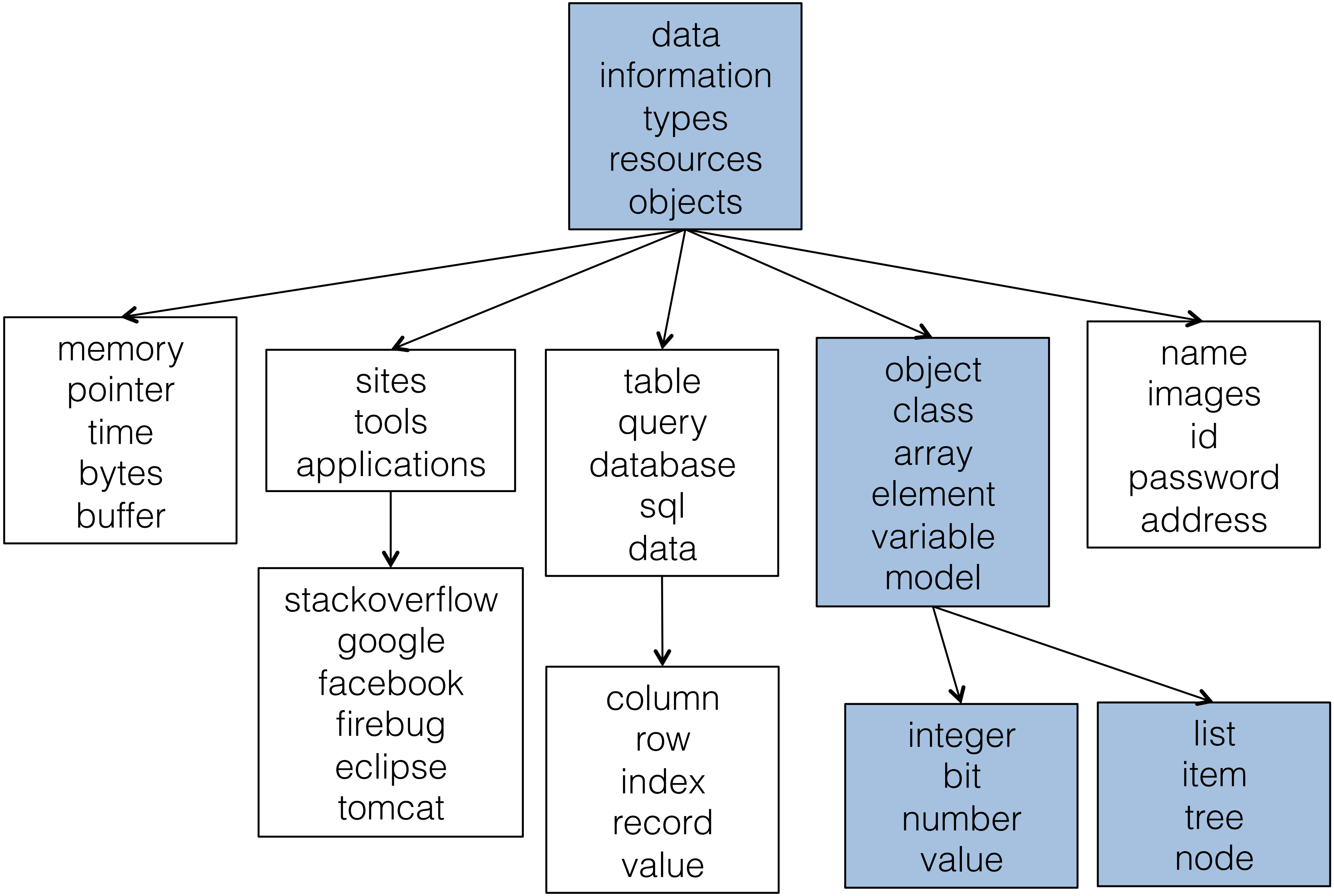
column
row
index
record
value

object
class
array
element
variable
model

integer
bit
number
value

list
item
tree
node

name
images
id
password
address



data
information
types
resources
objects

memory
pointer
time
bytes
buffer

sites
tools
applications

stackoverflow
google
facebook
firebug
eclipse
tomcat

table
query
database
sql
data

column
row
index
record
value

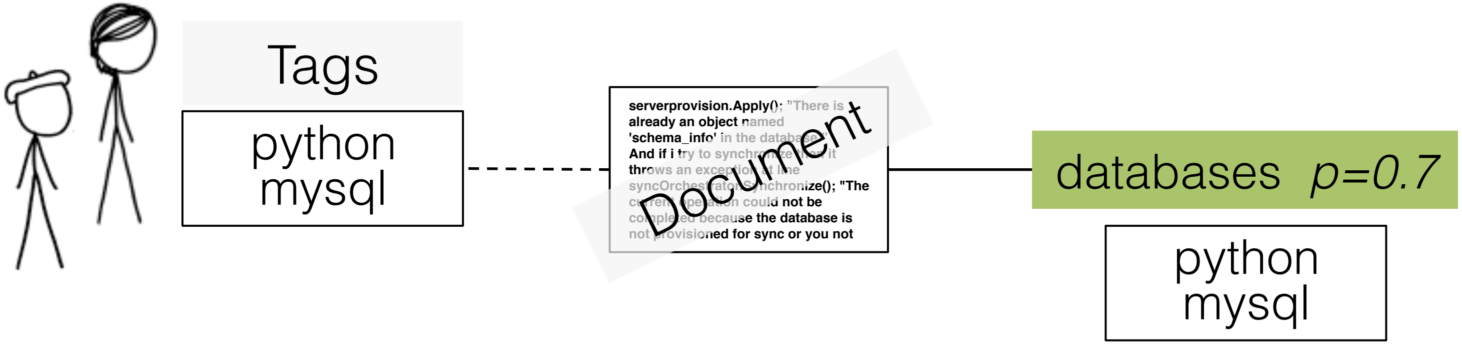
object
class
array
element
variable
model

integer
bit
number
value

list
item
tree
node

name
images
id
password
address

KB-LDA Topics versus Human-Provided Tags



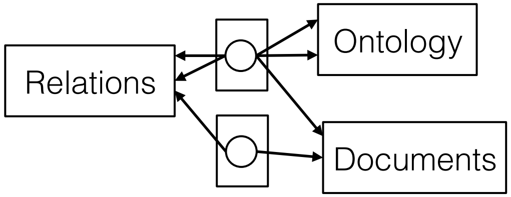
serverprovision.Apply(); "There is already an object named 'schema_info' in the database." And if i try to synchronize then it throws an exception at line syncOrchestrator.Synchronize(); "The current operation could not be completed because the database is not provisioned for sync or you not

For each document:

1. Find most probable topic (e.g., databases)
2. Aggregate tags for topic

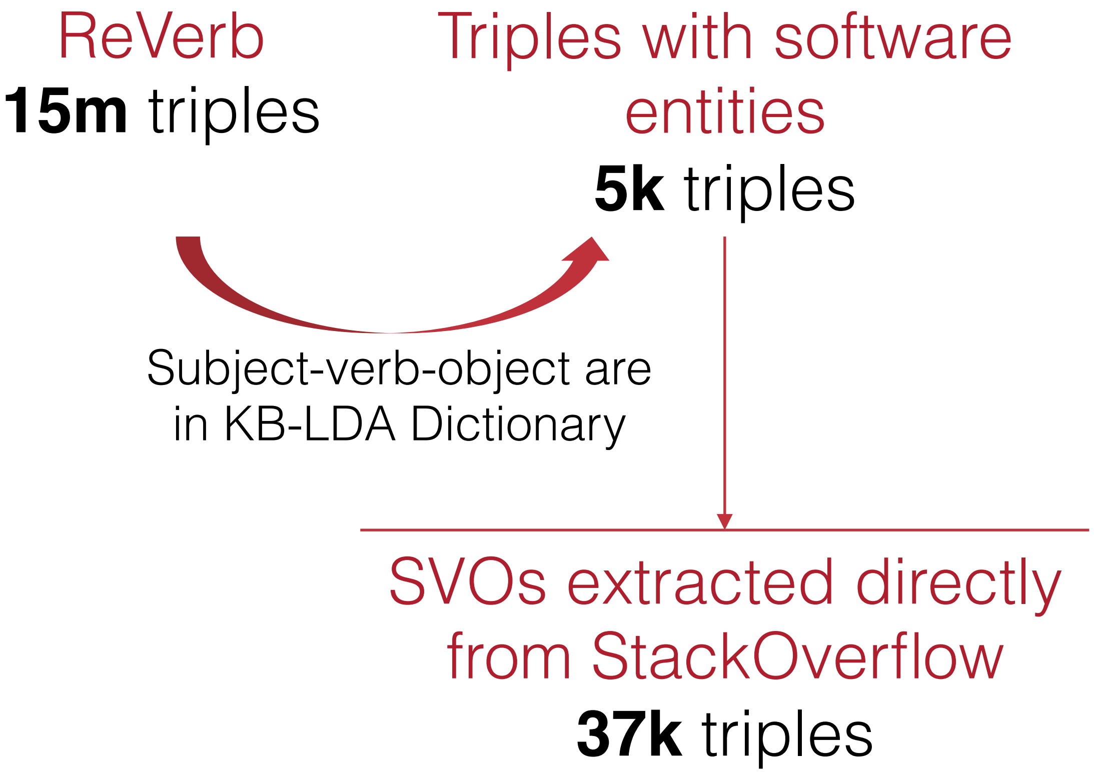
What tags are frequently associated with topics?

Top Tags Associated with Topics



	KB-LDA Topics	**Human Tags**
databases	table query **database** **sql** column	**sql** mysql **database** performance php
regular expressions	**string** character characters text line	regex **string** python php ruby
html elements	element div **css** elements http	**css** html jquery html5 javascript

Domain-Specific Extraction from Open IE



ReVerb
15m triples

Subject-verb-object are in KB-LDA Dictionary

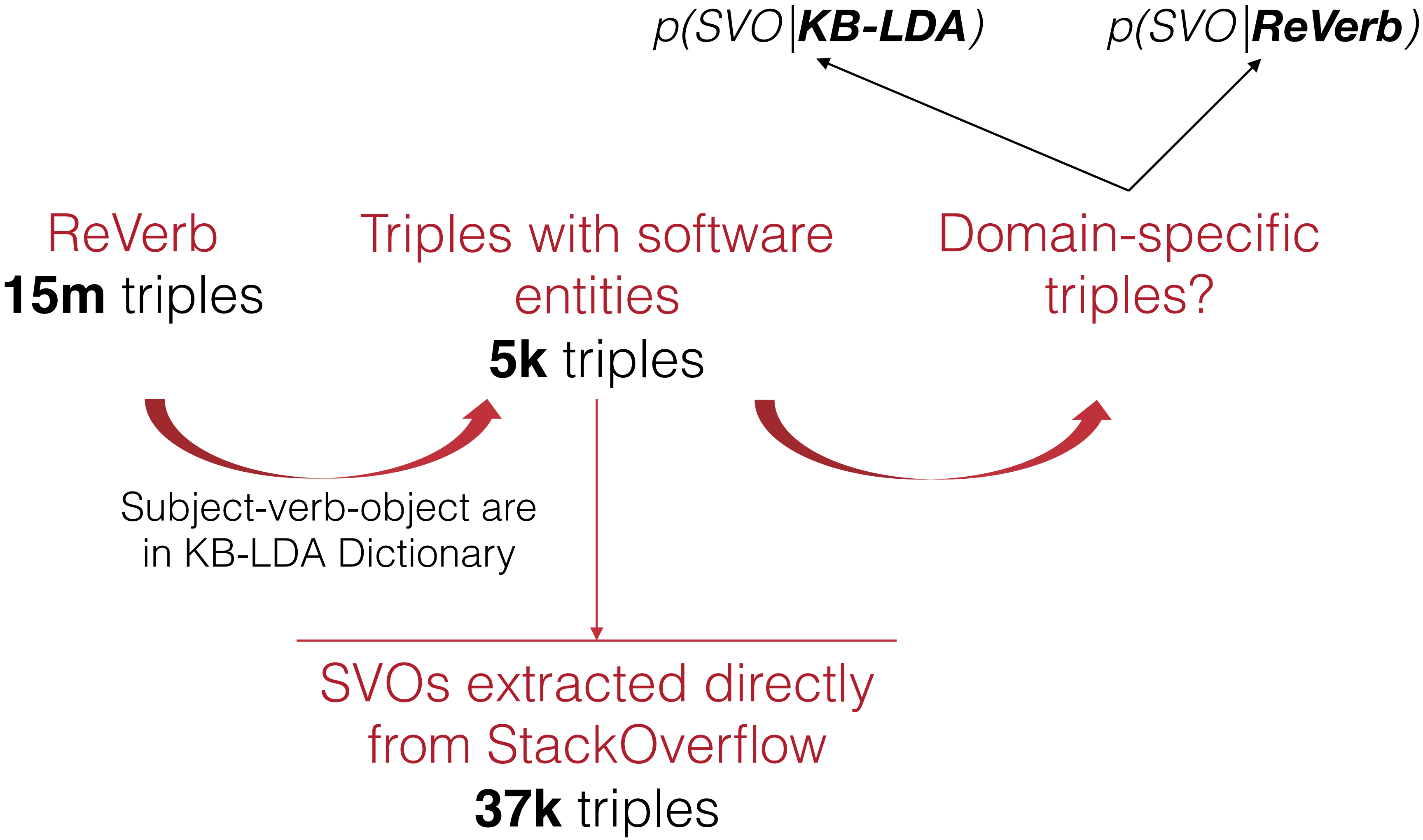
Triples with software entities
5k triples

SVOs extracted directly from StackOverflow
37k triples

✔ (safari, *supports*, svg)
✔ (computer, *is running*, xp)

✘ (view, *looks*, south)
✘ (people, *can read*, italian)

Domain-Specific Extraction from Open IE



$p(SVO|\boldsymbol{KB\text{-}LDA})$

$p(SVO|\boldsymbol{ReVerb})$

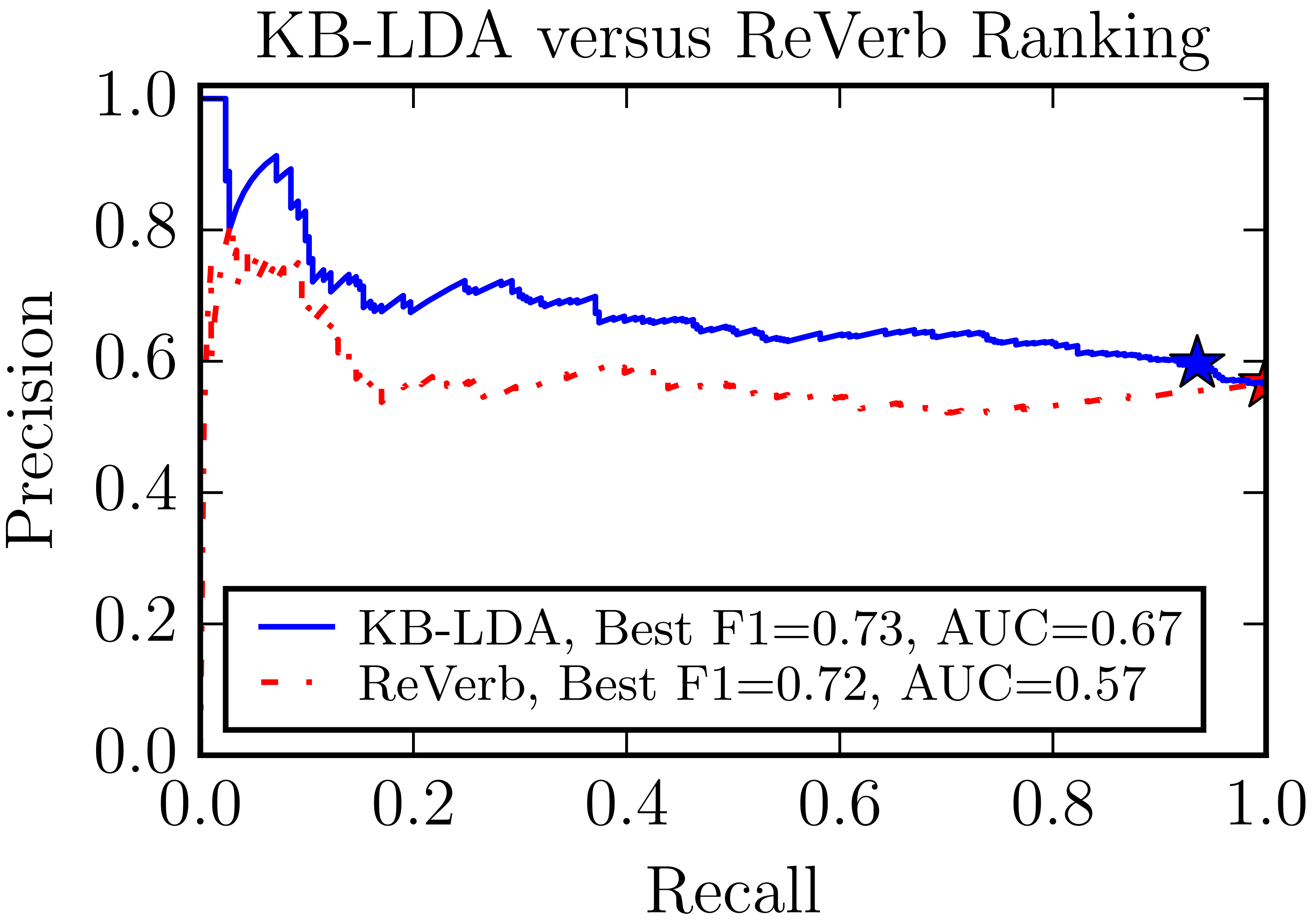
ReVerb
15m triples

Subject-verb-object are
in KB-LDA Dictionary

Triples with software entities
5k triples

Domain-specific triples?

SVOs extracted directly from StackOverflow
37k triples


KB-LDA versus ReVerb Ranking
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
KB-LDA, Best F1=0.73, AUC=0.67
ReVerb, Best F1=0.72, AUC=0.57

KB-LDA: Conclusion

- ✓ Corpus-driven KB construction: Jointly optimizes schema and facts

- ✓ Unsupervised: Useful for exploration of new domains

- ✓ KB-LDA can be used to extract domain-specific facts from an Open IE knowledge base

www.cs.cmu.edu/~dmovshov

Thank you!